Final Project Proposal

Song Genre Classification

Introduction A song genre classification (supervised) using 2 different methods: **Support Vector Machine** or **"Linear Regression" Method** for classification

Two potential data that can be used for this projects were retrieved from **Kaggle**

- **Dataset 1 (https://www.kaggle.com/datasets/priyamchoksi/spotify-dataset-114k-songs/data)**
  - This dataset contains +100k of samples which can be better used for the "linear regression" method
  - It includes several potential features, such as *duration_ms, explicit, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentallness, liveness, valence, and tempo*
- **Dataset 2 (https://www.kaggle.com/datasets/joebeachcapital/30000-spotify-songs)**
  - This dataset contains only ~30k of samples which is more suitable for the Support Vector Machine model
  - It also includes similar potential features as the first dataset, such as *duration_ms, explicit, danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentallness, liveness, valence, and tempo*

However, **dataset 2** seems to be more suitable as it enables to compare both model (either SVM or "linear regression" method) in terms of their accuracy and for learning purposes.

Dataset 1

In [32]:
```
import pandas as pd
import matplotlib.pyplot as plt

# for google colab (uncomment)
# from google.colab import drive
# drive.mount('/content/drive')
# filepath = '/content/drive/MyDrive/dataset1.csv'

df = pd.read_csv('dataset1.csv')
df.head()
```

Out[32]:

	Unnamed: 0	track_id	artists	album_name	track_name	popularity	duration_ms	explicit	danceability
0	0	5SuOikwiRyPMVoIQDJUgSV	Gen Hoshino	Comedy	Comedy	73	230666	False	0.676
1	1	4qPNDBW1i3p13qLCt0Ki3A	Ben Woodward	Ghost (Acoustic)	Ghost - Acoustic	55	149610	False	0.420
2	2	1iJBSr7s7jYXzM8EGcbK5b	Ingrid Michaelson;ZAYN	To Begin Again	To Begin Again	57	210826	False	0.438
3	3	6lfxq3CG4xtTiEg7opyCyx	Kina Grannis	Crazy Rich Asians (Original Motion Picture Sou...	Can't Help Falling In Love	71	201933	False	0.266
4	4	5vjLSffimiIP26QG5WcN2K	Chord Overstreet	Hold On	Hold On	82	198853	False	0.618

5 rows × 21 columns

Dataset 2

In [34]:
```
#for the second dataset
df = pd.read_csv('dataset2.csv')
df.head()
```

Out[34]:

	track_id	track_name	track_artist	track_popularity	track_album_id	track_album_name	track_albu
0	6f807x0ima9a1j3VPbc7VN	I Don't Care (with Justin Bieber) - Loud Luxur...	Ed Sheeran	66	2oCs0DGTsRO98Gh5ZSl2Cx	I Don't Care (with Justin Bieber) [Loud Luxury...	
1	0r7CVbZTWZgbTCYdfa2P31	Memories - Dillon Francis Remix	Maroon 5	67	63rPSO264uRjW1X5E6cWv6	Memories (Dillon Francis Remix)	
2	1z1Hg7Vb0AhHDiEmnDE79l	All the Time - Don Diablo Remix	Zara Larsson	70	1HoSmj2eLcsrR0vE9gThr4	All the Time (Don Diablo Remix)	
3	75FpbthrwQmzHlBJLuGdC7	Call You Mine - Keanu Silva Remix	The Chainsmokers	60	1nqYsOef1yKKuGOVchbsk6	Call You Mine - The Remixes	
4	1e8PAfcKUYoKkxPhrHqw4x	Someone You Loved - Future Humans Remix	Lewis Capaldi	69	7m7vv9wlQ4i0LFuJiE2zsQ	Someone You Loved (Future Humans Remix)	

5 rows × 23 columns