**Spotify Data Proposal:**

**Team Members:**

Musab

Ifa

Kaarjan

Rohan

**Project Description:**

The aim of this project is to analyse the dataset containing audio statistics of the top 2000 tracks on Spotify from 2000-2019. Through data exploration and analysis, we will investigate the relationship between various musical traits (such as danceability, speechiness, energy, valence, and tempo) and the popularity of songs. Additionally, we will explore artist characteristics, including track count, longevity, and genre, and their correlation with song popularity.

**Research Questions:**

Question 1:

We will investigate the correlation between musical traits (such as danceability, speechiness, energy, valence, and tempo) and the popularity of songs in the dataset. Through correlation analysis, I will uncover the associations between these attributes and the level of popularity. Additionally, I will identify which specific attributes exhibit the strongest correlation with song popularity.

Question 2/3:

Which artists have the highest number of tracks and the highest popularity scores in the dataset? Can we identify any correlations between artist characteristics (Genre, Longevity, Track Count) and the popularity of their songs?

*Bonus (if we have time):*

*Analyse the dataset and track the popularity of genres over time. By identifying the most popular genre for each year and plotting it on a graph, we will aim to visualize the evolution of genre preferences, uncover emerging genres, identify declining genres, and observe shifts in musical tastes across different decades.*

**Datasets to be used:**

We will be using this single dataset from Kaggle

<https://www.kaggle.com/datasets/paradisejoy/top-hits-spotify-from-20002019>

**Breakdown of Task:**

Task 1: Investigating the correlation between musical traits and song popularity

Task 2: Analyzing artist characteristics and their correlation with song popularity