

Data Analytics Project - Proposal

STAT 420, Summer 2023 - Soumya Nanda, Noam Isachar, Jonas Jansen

1. Team Members:

- Soumya Nanda - srnanda2
- Noam Isachar - noami2
- Jonas Jansen - jonasj2

2. Project Title:

Predicting Song Popularity

3. Data File Description:

The dataset is called **Spotify_1Million_Tracks** and it contains 1.47M observations (songs) and 20 variables. We will try predicting the song's popularity which is determined by several factors, including the number of streams, user engagement, and how recently the song was released. The dataset also includes other details about each songs like the song's name and artist, and a few metrics from Spotify such as the danceability, energy and acousticness (numerical) as well as the key, time signature, year and genre (categorical).

4. Data Background Information

The dataset's source is here: <https://www.kaggle.com/datasets/amitanshjoshi/spotify-1million-tracks>.

The data was extracted from the Spotify platform using the Python library "Spotipy", which allows users to access music data provided via APIs. It includes over 1 million tracks with 19 features between 2000 and 2023. Also, there is a total of 61,445 unique artists and 82 genres in the data.

5. Team Interest

The questions we would like to answer is whether a song's popularity could be predicted, potentially even prior to it's released, based on the song's musical characteristics and technical features. Is it possible to point out what can make a song more popular?

6. Data in R evidence

```
data <- read.csv("data/spotify_data.csv")  
  
head(data)
```

```

##  X   artist_name      track_name      track_id popularity year
## 1 0   Jason Mraz  I Won't Give Up 53QF56cjZA9RTuuMZDrSA6      68 2012
## 2 1   Jason Mraz 93 Million Miles 1s8tP3jP4GZcyHDsjvw218      50 2012
## 3 2 Joshua Hyslop Do Not Let Me Go 7BRCa8MPiyuvr2VU309WOF      57 2012
## 4 3   Boyce Avenue      Fast Car 63wsZUhUZLlh10syrZq7sz      58 2012
## 5 4   Andrew Belle Sky's Still Blue 6nXIYClvJAfi6ujLiKqEq8      54 2012
## 6 5 Chris Smither  What They Say 24NvptbNKGs6sPy1Vh100v      48 2012
##      genre danceability energy key loudness mode speechiness acousticness
## 1 acoustic      0.483  0.303  4  -10.058  1      0.0429      0.6940
## 2 acoustic      0.572  0.454  3  -10.286  1      0.0258      0.4770
## 3 acoustic      0.409  0.234  3  -13.711  1      0.0323      0.3380
## 4 acoustic      0.392  0.251 10   -9.845  1      0.0363      0.8070
## 5 acoustic      0.430  0.791  6   -5.419  0      0.0302      0.0726
## 6 acoustic      0.566  0.570  2   -6.420  1      0.0329      0.6880
##      instrumentalness liveness valence tempo duration_ms time_signature
## 1      0.00e+00    0.1150  0.139 133.4      240166      3
## 2      1.37e-05    0.0974  0.515 140.2      216387      4
## 3      5.00e-05    0.0895  0.145 139.8      158960      4
## 4      0.00e+00    0.0797  0.508 205.0      304293      4
## 5      1.93e-02    0.1100  0.217 171.9      244320      4
## 6      1.73e-06    0.0943  0.960  83.4      166240      4

```