**RE: Popular Music**

[**Link to dataset**](https://drive.google.com/file/d/1LjOhTjreCxwJU5TUkDbMMk10NN9lBL45/view?usp=sharing)

**Context**

You’ve been asked by a client to analyze a data set and give recommendations based on the core questions below. This client is a new music venue serving a diverse international audience. They want to understand what music is popular to their prospective patrons, both currently and in the future. They also want to understand what promotions or events to have to encourage attendance to the venue.

They’ve obtained a dataset with the top songs for 2019 across the world, Africa, and in various countries that their patrons are from. Using this data they’d like you to give them a presentation and set of insights and recommendations to help them select their music and drive customer traffic.

**Core Questions**

1. What are some of the basic trends and summaries of the data?
2. What features are associated with song popularity? Do those features vary by location?
3. Which countries have similar tastes in music?
4. Can we predict how popular a song (not in this data set) will be?
5. Can the songs in this dataset be clustered into particular groups or segments?
6. What kind of “themed” nights or promotions can we have to encourage attendance?

**Dataset Detail**

This data contains the most popular songs from various countries around the world. Note that songs can appear on this list multiple times (e.g. if it’s a popular song in multiple countries).

The variables are as follows:

* Title, Artist, Genre: Self explanatory
* Release: Year of release
* Beats Per Minute: The tempo of the song
* Energy: Score on how energetic the song is
* Danceability: Score on how danceable the song is
* dB: Score on how loud the song is
* Live Performance Score: A rating on whether the song was likely recorded live (higher means more likely the song is recorded live)
* Positivity: Positive or negative valence of the song
* Song Duration: Length of song (in seconds)
* Acoustic: Rating of how acoustic the song is
* Speech: Rating of how much the song is focused on spoken word
* Popularity: Rating of how popular the song is
* Country: The country where that song is popular

After answering the core questions, please give an overview document addressing the following points:

* 1. **Context and Data Audit** – Details on the specifics of the dataset, the primary summary statistics of interest, and any limitations to generalizability for the partner.
  2. **New Measures Constructed** – What new measures did you construct from the dataset?
  3. **Segmentation and Groups of Interest** – What natural groups are emerging in the dataset, and how can we connect these to a relatable population for the partner? What are the primary behavioral / activity differences between any groups proposed?
  4. **Potential Behavioral Insights** – Where are potential behavioral biases observable in the dataset?
  5. **Areas for Further Analysis** - Your time will be limited, and we don’t expect everyone to produce the most complex models possible, so propose what you would do if you had 4 more days on the dataset, and why you feel it is important.
  6. **Proposed New Measures** – Beyond new analysis, there are likely also measures that would have contributed significantly to the analysis provided they were in the data collection tool. Share any primary metrics of interest, as well as your rationale for inclusion.