## Project Topic Idea: <Music Dashboard*>*

Of the project ideas you submitted in the previous deliverable, select the one you want to work on throughout this project. In making your decision, please refer to the feedback you have on the previous deliverable. If all ideas were found plausible, feel free to select the one that you liked the best **as a team**.

Complete this section based on the previous deliverable. After deciding on your topic idea, simply copy and paste same information from the Topic Proposals document here.

### Problem Statement

* There is no standard way to rank and identify trending genres, albums, and artists. The availability of vast amounts of user data on music-sharing platforms is increasing competitiveness between music industry companies, especially when signing artists and promoting appropriate content

### Significance of the Problem

* + Significance
    - Streaming services have increased music companies’ pressure to utilize big data/data science to analyze user behavior to adapt strategies.
    - Music is a huge industry. According to Forbes, the global music industry sales rose by 10% in 2018 to $19 billion. There is a lot of money to be made through satisfying market demands.<https://www.forbes.com/sites/hughmcintyre/2019/04/02/the-global-music-industry-hits-19-billion-in-sales-in-2018-jumping-by-almost-10/#174acd7a18a9>
    - Music is important to our society. It shapes our emotions, perception of life, and culture.
  + How could our insights be useful?
    - Streaming companies need to recommend appropriate songs to ensure user retention and satisfaction. They can adapt their algorithm based on our findings on genre, album, song, artist, etc. trends.
    - There is a big gap between music that is consumed vs. created. This dashboard can also allow budding artists to understand the market demands and appropriate channels to promote.

### Dataset(s)

* Link to datasets: <https://medium.com/@aaron.turnbull76/exploring-spotifys-top-200-for-2018-b6fb1cb7b3de>
* Song\_features: audio features, like tempo and liveliness
* Artist\_info: information about the artists, like followers and popularity
* Streams: number of streams for each top 200 song for 2018

## Dataset File

Download or scrape your data from the source you identified above. Save your dataset as a CSV file. The first row of the file should contain variable names.

Describe your variables below (add more rows if necessary):

|  |  |
| --- | --- |
| **Variable name in file** | **Description** |
| Position | Ranking within the Spotify Top 200 List |
| Track Name | Name of song |
| Artist | Artist of song |
| Streams | Number of streams |
| URL | Link to song on Spotify |
| Date | Date song made Top 200 |
| Region | Location tied to song |
| Track\_id | Spotify’s identification number for the song |
| Artist\_id | Spotify’s identification number for the artist |
| Followers | Number of followers on Spotify for the artist |
| Genres | Genre of song |
| Popularity | Popularity Ranking |
| Energy | From 0 – 1, measure of intensity and activity, ranked by Spotify |
| Liveness | Presence of audience in recording. .8 and above = strong likelihood |
| Tempo | Beats per minute tempo |
| Speechiness | Presence of spoken words. .66 and above likely made entirely of spoken words. .33 - .66 may have both. |
| Acousticness | 0-1 confidence measure of whether track is acoustic, with 1 as high. |
| Instrumentalness | Predicts whether track has no vocals (.5 and above using instrumental, 1 is highest) |
| Danceability | How suitable track is for dancing based on tempo, rhythm stability, beat strength, regularity (1 = most danceable) |
| Key | Overall key of song (-1 means no key detected) |
| Duration\_ms | Duration of track in milliseconds |
| Loudness | Loudness in decibels, typically from -60 to 0 db |
| Mode | Modality (major 1 or minor 0) |
| Valence | Measure from 0 – 1 of positiveness (cheerfulness) |

Submit a CSV file, or multiple files, containing your data.

If you have scraped your data, you should also submit a Jupyter Notebook containing your Python code used to scrape the data. Please be reasonable and comment your code out whenever it makes sense to do so.