Cogs 260 Mini-Project Proposal
Lead Author: Morgan Fitzgerald

Co-authors: Michaela Cullum-Doyle, Vanessa Martin, Sowmya Manojna

Title: The Impact of the Covid-19 Pandemic on Music Preferences: A Study on Behavior Modification

Research question.

Music has become an integral part of our daily lives, facilitated by technological advancements that allow easy access to a vast range of music genres, leading to changes in music preferences. However, the Covid-19 pandemic has significantly impacted our behavior and emotions, sparking interest in studying music preferences before and after the pandemic. This study aims to analyze trends in music preferences during the period of 2017 to 2022 and assess how technology and the pandemic have influenced music preferences over time.

Given the significant impact of music on mental health and well-being (Lin et al., 2011), it is crucial to understand how music has influenced people's lives during the pandemic. Examining changes in music preferences during the Covid-19 pandemic can provide valuable information on alterations in mood, emotional state, and overall psychological health, offering insights into how the pandemic has affected people's behavior and emotions. Moreover, these changes can be useful in the music industry, marketing, and advertising, and can aid the development of public health and mental health interventions during future public health crises.

Data/materials. We will use a dataset from Spotify Billboard Year-End top 100 Spotify songs from 2017 to 2022 with 13 song attributes. This dataset was scraped from Spotify. The dataset contains 18 columns (14 features, song title, artist, ranking and spotify song ID) and 600 rows of data. In order to analyze trends in music preference that have been affected by the COVID-19 Pandemic, we used the Billboard Year-End Top 100 songs. The top 100 songs for each year from 2017 to 2022 was scraped from Spotify. The audio features of each of these songs were then queried using the Spotify API. A total of 14 audio features were obtained for each song. The audio features are:

- Streams (quantitative)
- Danceability (quantitative)
- Energy (quantitative)
- Key (categorical)
- Loudness (quantitative)
- Mode (categorical)
- Speechiness (quantitative)
- Acousticness (quantitative)
- Instrumentalness (quantitative)
- Liveness (quantitative)
- Valence (quantitative)

- Tempo (quantitative)
- Time Signature (categorical)

File link	Format	Description
https://drive.google.com/drive/folders/1gaSWuWqmfvfTVlOqf5qoXBStbGXkmvSQ?usp=share_link	CSV	6 CSV files, one each for the Billboard year-end top 100 songs from years 2017 to 2022. Each row corresponds to a single song, with a unique title (first column). The remaining columns contain the song attributes, i.e. artists, genre, year, beats, energy, danceability, loudness, liveness, valence, length, acousticness, speechiness, and popularity.

Course impact/relevance. This dataset enables projects within the domains of linear regression, hypothesis testing, prediction and model evaluation among others.

Outcome(s).

1. Hypothesis testing: The project aims to test the hypothesis that the COVID-19 pandemic has altered music preferences. To do this, we will analyze the dataset of the top 50 Spotify songs from 2010 to 2019 and compare it to the music preferences before the pandemic. We will use statistical hypothesis testing techniques to determine if there is a significant difference between the two periods.
2. Linear regression: The project may also utilize linear regression techniques to predict future music popularity based on previous song preferences. The project will use the 13 quantitative song attributes available in the dataset, such as beats per minute, energy, danceability, loudness, liveness, valence, length, acousticness, speechiness, and popularity, to develop a predictive model. The model will analyze the relationship between these attributes and the popularity of the songs to predict future trends.
3. Dimensionality reduction: Dimensionality reduction techniques can be applied to the quantitative song attributes available in the dataset, such as beats per minute, energy, danceability, loudness, liveness, valence, length, acousticness, and speechiness, to identify the most important attributes that contribute to music popularity.
4. Model evaluation: The project will evaluate the predictive model's performance using various metrics such as mean squared error (MSE).
5. Unsupervised Learning: Unsupervised clustering and manifold learning approaches can be used to analyze the data, identify clusters of songs that share similar features. This approach would enable us to suggest new songs based on an individual's listening history.

References:

Lin, S. T., Yang, P., Lai, C. Y., Su, Y. Y., Yeh, Y. C., Huang, M. F., & Chen, C. C. (2011). Mental health implications of music: Insight from neuroscientific and clinical studies. *Harvard review of psychiatry*, *19*(1), 34-46.