ECON 312 FINAL PROJECTS ROUGH IMPLEMENTATIONS

Members: Nguyen Tran, Tobi Bui, Audrey Burkey

Project Description:

The objective of our project is to collect and visualize data on the songs from the Billboard Hot 100 list between 2010 and 2020 and build a simple prediction model using Linear Regression. We will use that model to pick out the top 10 songs of 2021 and 2022 with the highest popularity prediction. We also build a simple recommendation system based on mood and also a filter for artists, years and genres.

Top 30 artists between 2010-2020:

- This graph visualizes the number of songs on the Billboard Hot 100 list the top 30 artists had between 2010-2020. Drake leads the charts with more than 25 songs.

Top 10 genres between 2010-2020:

- Four out the ten top genres of the Billboard Hot 100 lists between 2010 and 2020 include elements of “pop” music, with generic pop music having the most hits. Pop music is defined as music that serves purposes for “light entertainment, commercial imperatives, and personal identification.” These associations with pop music make radio and subsequently the Billboard Hot 100 great environments for pop music to thrive.

Distribution of major or minor keys in hit songs between 2010-2020:

- There are more songs in a minor key than a major key within our dataset, showing that songs that are less positive, happy or peppy are more popular. This coincides with the data showing that the valence of hit songs has decreased over time.

Time signature distribution between 2010-2020:

- 95.2% of songs on the Billboard Hot 100 list between 2010 and 2020 used the time signature of 4/4, a simple measure mostly used in pop songs.

Correlation Matrix:

- The correlation matrix displays the correlation values between the variables we are analyzing between these songs. This matrix gives us insight into what factors contribute to a song being on the Billboard Hot 100 list between 2010 and 2020. There is a medium correlation of .36 between an artist’s popularity and their track’s popularity, illustrating that if an artist is already perceived to be popular, then their track will perform better. Most other variables had a weak correlation with a track’s popularity, but “acousticness” and “danceability” both stick out as good qualities to have in a hit song.

Scatterplot of track duration in milliseconds and track popularity.

- There is no clear correlation between track length and track popularity.

Correlation between danceability and valence (song mood)

- The higher the song mood, the more it is perceived to be danceable.

Correlation between tempo, energy and popularity

- There is a positive correlation between track popularity and track tempo, leaning towards a higher tempo for a more popular song.

Tracking Danceability, energy, and valence by years:

- We see an overall incline in the amount of “danceability” a song has on the Billboard Hot 100 between 2010 and 2020. The valence (described as the musical positiveness of a song) of songs has gone through many trends but overall we have seen a decrease in the positiveness of songs between 2010 and 2020. Finally, the energy of songs has decreased, surprisingly, even when danceability of songs has increased.

Tracking speechiness, acousticness, and liveness by year

- Out of the variables shown, we see that speechiness has increased the most between 2008 and 2018, followed by the acousticness of a song. The increases in these variables coincide with the increase in “danceability” of songs, leading to more instrumental-leading hit songs rather than speech or lyric leading hits.

Conclusions:

Based on our visualization, we have concluded the top three variables of importance when concocting a song to be featured on the Billboard Hot 100 list between 2010 and 2020: danceability, valence, and speechiness. Danceability and valence seem to go hand in hand, creating an upbeat, happy / positive track. The increase in speechiness can be thought of through the framework of rap and spoken words being called out through a song, again correlating to more danceable songs. In conclusion, our data shows that the most popular songs between 2010 and 2020 were danceable and upbeat, with callouts throughout, making the perfect club and radio hit to be featured on the Billboard Hot 100 list. These popularity trends also correlate to an increase in popularity of video-making apps like Tik Tok; some of the most popular videos on the app are dance videos, possibly giving incentives to artists to create more “danceable” tracks.

2021 and 2022 popularity predictions:

We have built a simple Linear Regression model through scikit-learn, using these following variables: 'danceability','energy','loudness','speechiness','acousticness','liveness','valence','tempo', 'artist_pop'. We decided to remove keys, mode and instrumentalness since they show little correlation to the output popularity we are predicting. We have outputed our own top 10 charts for those two years.

Deploying on a web using Streamlit (80% progress):

We have used streamlit to deploy our graph and recommendation system onto it. We made the user choose between moods (i.e. happy, sad, energetic) and gave out some songs for them. Users can also filter by artists, years (using sliders) and genres. It is still not fully completed, since we are examining some more ways to filter the mood out efficiently, but right now we have already got some output for users.