sabinabains / **dsc-capstone**   Public

<> **Code**   ⊙ Issues   ⅋ Pull requests   ▷ Actions   ⊞ Projects   📖 Wiki   ⊘ Security   ∿

ᛈ main ▾

dsc-capstone / **README.md**

🅐 **sabinabains** fixed typos and included recommendations   🕓 **History**

👥 **1 contributor**

≣   78 lines (48 sloc)   8.28 KB   ···

PREDICTING TRACK SUCCESS ON SPOTIFY USING COVER ART

A picture is said to be worth a thousand words, but does that saying still apply to album art? Before streaming services such as Spotify became the preferred format for music consumption, listeners would peruse the aisles of music stores, scanning the album artwork of CD's and Vinyl Records. During this time, album art likely played a role in whether the album was purchased or not. Our goal is to predict whether this still is true, and if so, determine which album covers are predicted to be most sucessful.
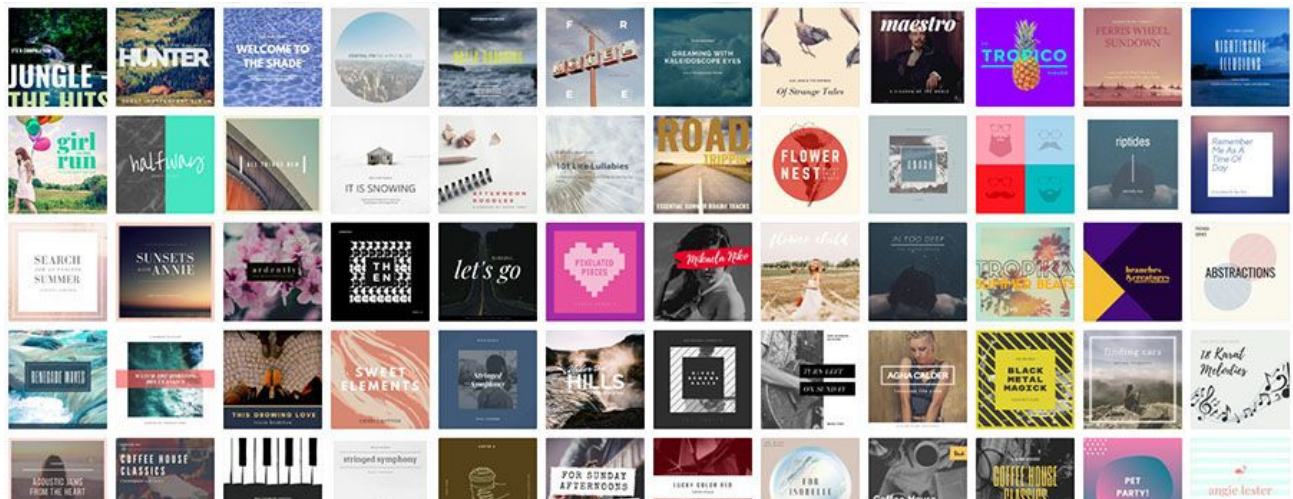
Business Understanding

Album art has played an important role in music. It gives visual representation and additional context to the story behind an album. Perhaps most importantly, it helped artists sell music. Before streaming was introduced into the world of music, albums were largely judged by their artwork. Iconic albums such as "Abbey Road" by The Beatles and "Nevermind" by Nirvana have artwork that are still frequently talked about today.

However, in recent years streaming services such as Spotify have dominated, with streaming making up 80% of revenue in the U.S. music industry. Artists and labels have been left wondering if this shift from hard copies to streaming has affected the prominence of albums, and therefore album artwork. The Playlist Effect is a phenomenon that suggests with the rise of streaming, subscribers largest listen to curated auto-play playlists with individual tracks rather an albums as a whole.

with Spotify leading the music streaming industry, it's crucial for record labels to understand the effectiveness of albums and their artwork in this space. This analysis will look at the popularity of tracks from Spotify's discovery playlists and their respective artwork to determine which album styles are associated with track success.
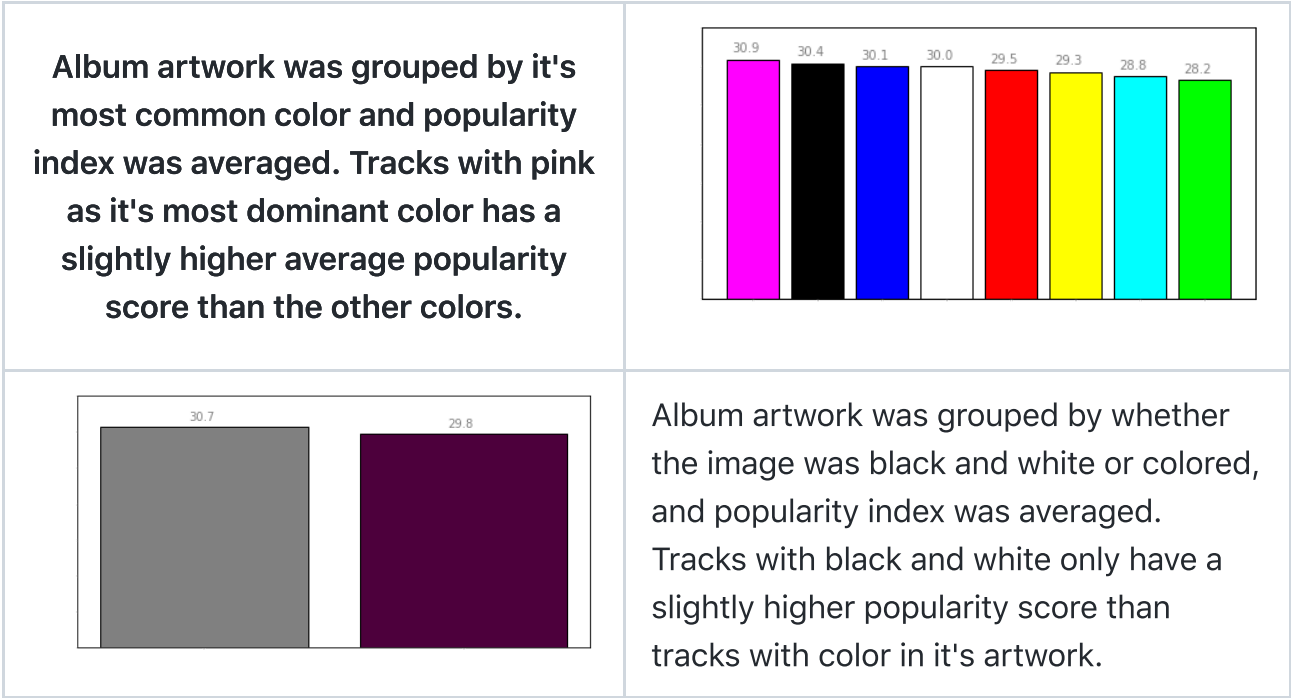
Data Understanding

Data for this project was pulled from the Spotify API SpotiPy, however the data pulled for this analysis has also been saved in the data folder of this repo.

The input variables are 1,835 album artworks from tracks in Spotify's "Fresh Finds" playlists. These playlists are solely comprised of independent artists and labels to eliminate potential outliers from hugely sucessful artists who will likely generate many streams regardless of album art. The following genres/groups were included: Korea, Vietnam, Hip-Hop, Indie, Brasil, 2021, Folk, NZ, 2019, IE, Country, Dance, Latin, R&B, Rock, GSA, Jazz, Pop, Experimental, Italia, España, EDM, All Genres.

the target variable is the track's popularity index, an integer between 0 and 100, with 100 being most popular. This is defined as "The popularity is calculated by algorithm and is based, in the most part, on the total number of plays the track has had and how recent those plays are. Generally speaking, songs that are being played a lot now will have a higher popularity than songs that were played a lot in the past."

Preliminary analyses on album colors and average popularity do not show any correlation.

Album artwork was grouped by it's most common color and popularity index was averaged. Tracks with pink as it's most dominant color has a slightly higher average popularity score than the other colors.

Album artwork was grouped by whether the image was black and white or colored, and popularity index was averaged. Tracks with black and white only have a slightly higher popularity score than tracks with color in it's artwork.

Modeling and Evaluation

For model input each album artwork was imported and resized to a 3 dimensional shape of (60, 60, 3), where each value in the height and width represents a pixel of the image, and the depth represents the R, G, B color values of the image.
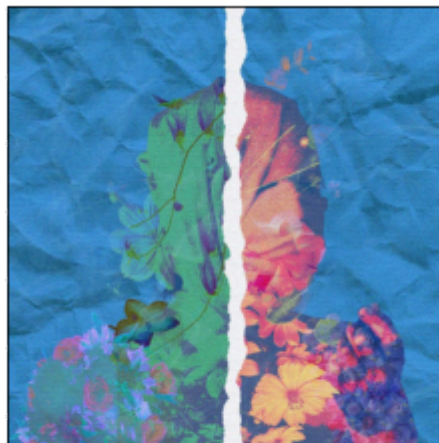
Convolutional Neural Networks (CNN's) were chosen for this Image Classification Regression analysis. CNN layers are able to reduce the high dimensionality of images without losing its information, which will allow the computer to "view" each image.

The first model run kept the initial dimensions of Spotify album artwork, (664, 664, 3). with one CNN layer and pooling layer. This model did not perform well, as it used too much computational power.

After resizing photos to (60, 60, 3) the models ran much faster. The remaining models tested involved adding / removing layers, increasing epochs, and tuning hyperparameters to find the lowest MSE score. The final model chosen generated an MSE of 140, with a filter size of 3 x 3, a pooling size of 2 x 2, linear activation, and no padding. Although this model outperformed the others, it is not an effective predictor for Spotify's popularity index. The scatterplot of true vs. predicted values demonstrate this, as well as it's R Squared score of 0. This score essentially tells us the model is about as efficient at predicting popularity than a simple averaging of popularity, 30.1, would be able to predict.

Epoch by Mean Squared Error	Scatterplot of True Vs. Predicted Values
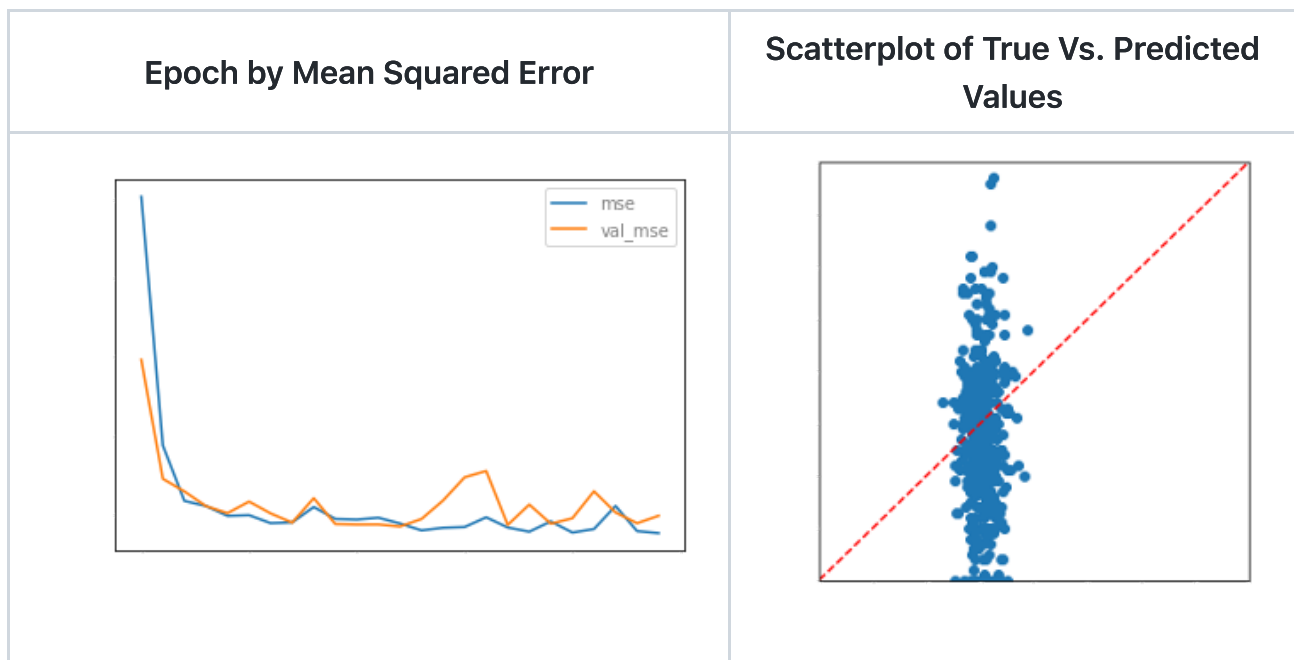
Conclusion

Due to the model's MSE and R Squared score, it is evident this model has poor predicitve ability and should not be used to gauge a track's popularity on Spotify. A lack of predictability could serve as an insight in itself however, as it is highly possible that users no longer focus on album artwork.

With more time I would attempt to support the above theory by running the same analysis on albums from the 80's up to the early 2000's. I would also look into finding a different target variable to measure "success". Spotify's metric is based on recency, which therefore overlooks older music. It also rates popularity comparatively to other songs, which is problem a because songs that may be popular within their own genre could be understated due to the popularity of other genres. Lastly, I would run a NLP analysis on track name, and see if this has any effect on popularity among Spotify listeners.

Based on the above results, I would make three recommendations to the client.

- Consider allocating resources away from album art creation, as artwork style doesn't seem to affect popularity on Spotify
- Commission a study on album and/or song name to understand naming effect on Spotify popularity
- If artwork context is crucial to telling the album story, consider other means of distribution outside of Spotify, such as physical stores, as users don't seem engaged with artwork on Spotify

Repository Navigation

Reproduction Instructions

To gather the same results as what's detailed in this .README, simply run the final_model.ipynb notebook. This will read in the data saved in this GitHub.

If the user would like to rerun this analysis with the most current Spotify playlists, The user needs to create a developer account and gain credentials for Spotify's API. Details are linked here. Once the user has a CID and Secret key, those should be pasted in the third cell of Spotify_API_pull.ipynb. This notebook can then be run, and the new album artwork and popularity indices will be saved. The user can then run final_model.ipynb to generate CNN models.

Repository Structure

```
├── PDFs : PDF files for submission
├── data : data used for modeling, includes album art and popularity index
pulled from Spotify API
├── images : images used in PPT and README
├── previous_files : older ppt and ipynb files
├── README.md : project information and repository structure
├── Spotify_API_pull.ipynb : notebook used to pull from API
├── cnn_models.ipynb : modeling, including the final model
├── helper_functions.ipynb : additional functions to support analyses
├── model_1.ipynb : first model run, separated due to large processing
times
└── predicting_song_popularity_by_album_artwork.pptx : Presentation for
stakeholders
```