PREDICTING THE POPULARITY OF SONGS ON SPOTIFY

BY

AAKASH KUMAR HARIT

AGENDA

- Motivation

- Project goals – Questions to explore

- Results

- Data Profile

- Summary and Future Recommendation

MOTIVATION

Can music be explained by science?

Is it possible to predict whether a song will be popular based on its any features?

- With the advent of the internet and advances in music technology, several new genres emerged after the 1990's diversifying the music industry.

- I challenged myself to find out if there is a scientific reason behind some songs being more popular than the others.
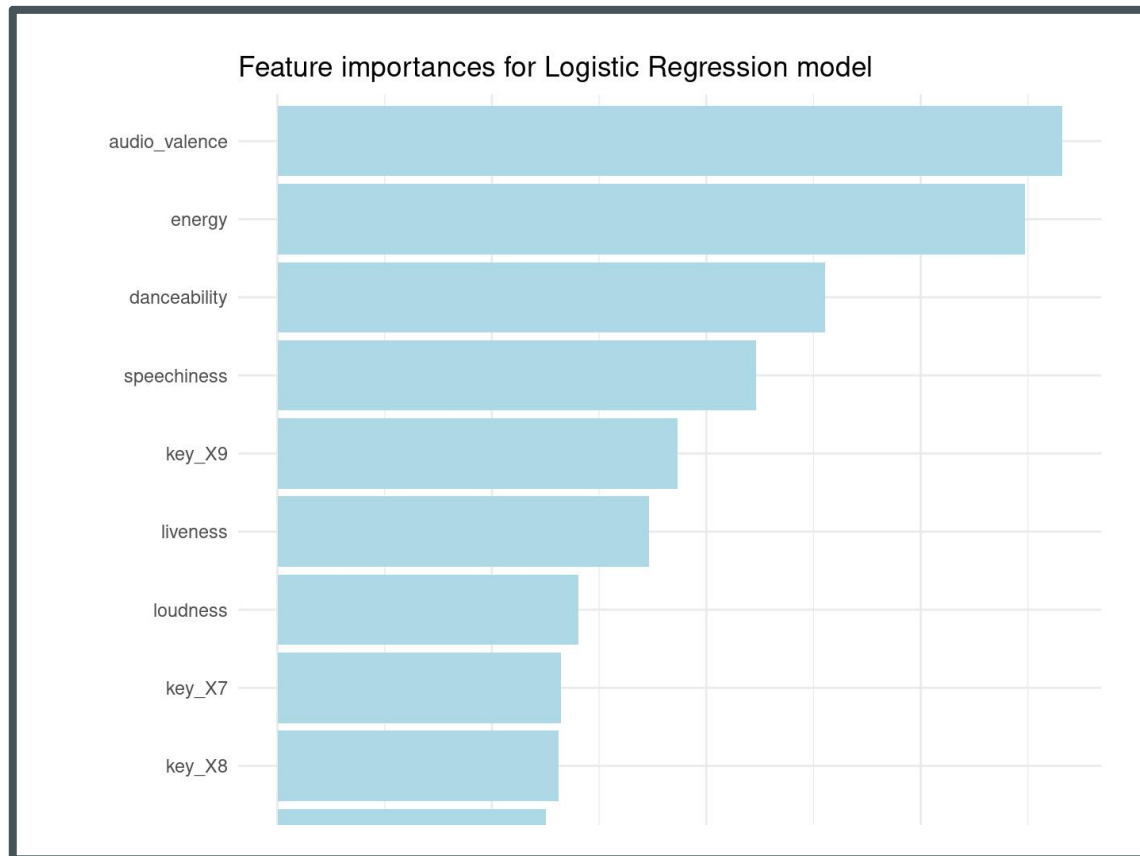
PROJECT GOALS

To understand features that appeal to a wide variety of listeners through data-driven outcomes.

Questions to explore:

1. Can we predict the popularity of a song just by its any features?

2. Which features contributes most to the popularity of a song?

3. What are the average features of a popular song, like duration tempo, energy etc?

RESULTS

Feature importances for Logistic Regression model

I ran 7 different models: Logistic , gaussian , Stochastic , KNN, decision tree, random forest, SVM.

Random forest worked best with cross validation F-1 score of 80.5%.

7 most important features:

- instrumentalness

- accousticness

- danceability

- energy

- speechiness

- loudness

- valence

RESULTS: DIFFERENT MODELS SCORES

Precision – Of the songs we classified as popular, how many were correct?

Recall – of the songs we should have classified as popular, how many did we?

F1 – weighted average of precision and recall

Efficiently measures correct classifications

MODEL	F-1 Score
Logistic Regression	75.3%
gaussian	75.4%
Stochastic gradient descent	74%
K – nearest neighbors	78.2%
Decision tree	77.5%
Random Forest	**80.5%**
Support Vector Machines	76.7%

VALUE ADDITION FROM MODELS

- Help music directors and producers to understand features that appeal to a wide variety of listers and optimize their revenue.

- Help songwriter and upcoming artists to write songs that would appeal to a larger audience , hence giving them a chance at popularity.

- Provide music composers and directors with a desired range of range of each individual feature that attributes to success of a song.

- Help music enthusiasts find popular music by recommending songs in spotify library based on popularity metric.

- Discover customer segments amongst Spotify users based on important song features for a given decade.

DATASET : DATA PROFILE

- Dataset consists of over 40k songs spread across the last 6 decades.

- Source: Kaggle (dataset pulled from Spotify API)

- Dataset has 16 features and a target variable (0 or1), where 1 tells song was popular.

- Some mentioned features: track , artist, danceability, energy, key , loudness, mode, speechiness , accoustiness , instrumentalness , liveness, valence, tempo etc.

- Definition of popular : the song was featured in weekly list of hot 100 tracks (By billboards) in that decade at least once.
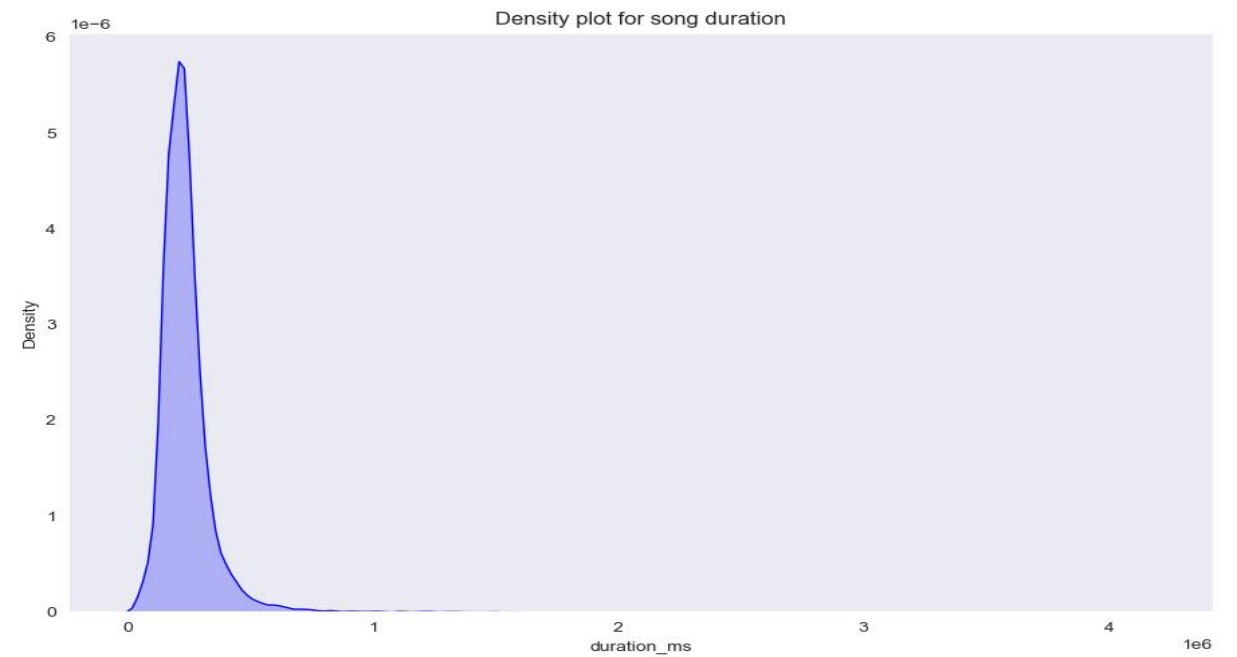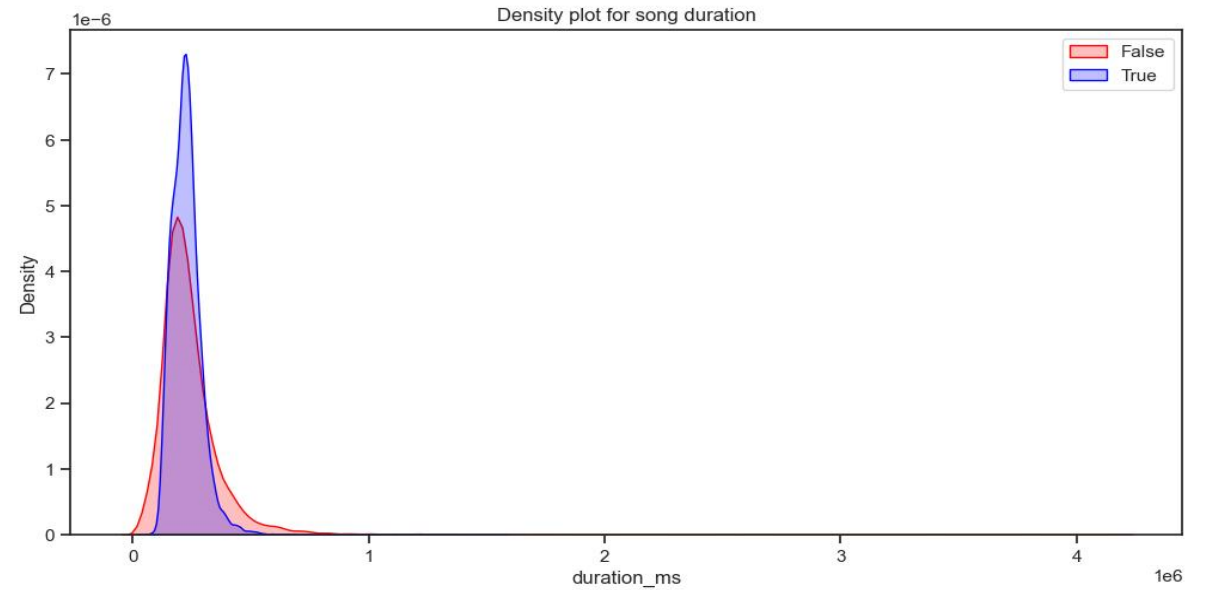
Multiple feature plots

data

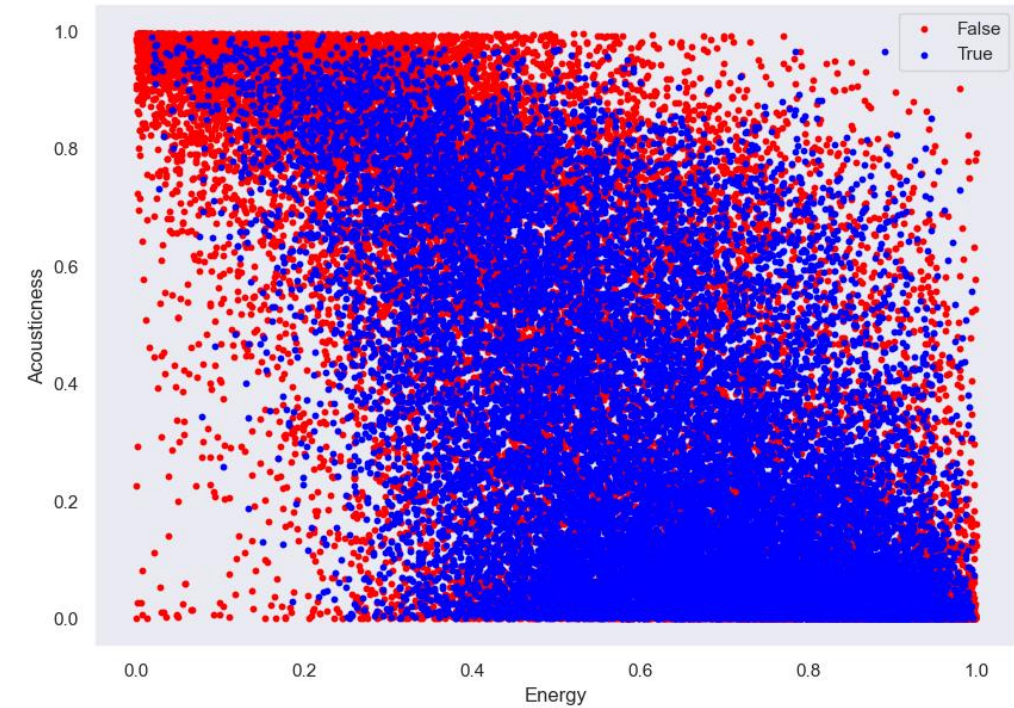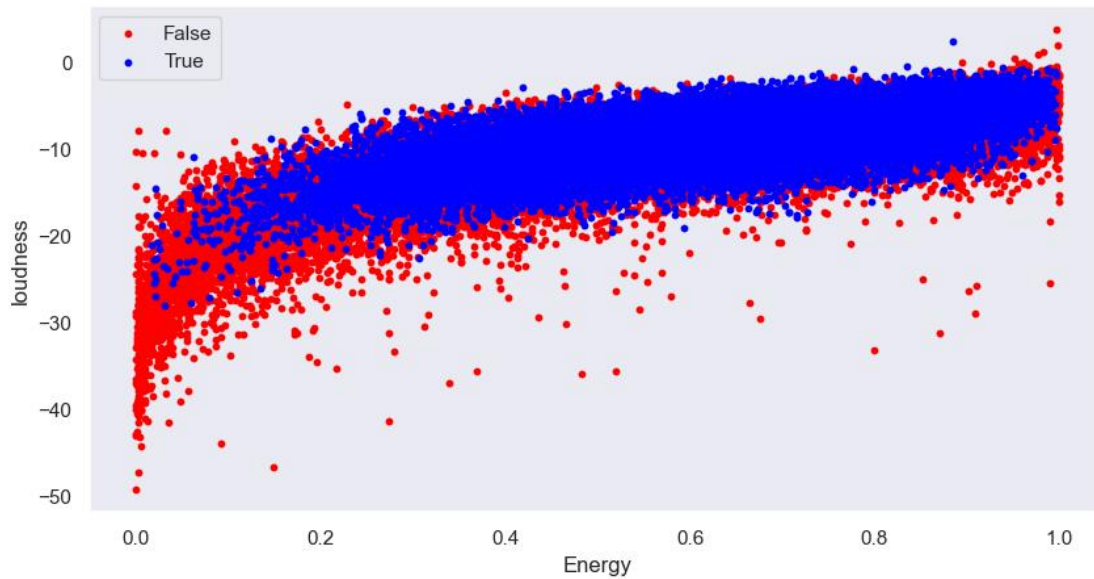|	track	artist	uri	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness
6]:	Attaining - Take 1 / Alternate Version	John Coltrane	spotify:track:3EwLV5hZqLKx5e0Lp1QcB7	0.342	0.462	4	-12.931	0	0.0389	0.51400	0.018100

liveness	valence	tempo	duration_ms	time_signature	chorus_hit	sections	target	decade
0.0977	0.397	98.466	816867	4	24.84938	33	0	1970

EXPLORATORY DATA ANALYSIS

TOTAL UNIQUE ARTIST : **11904**

UNIQUE TRACKS : **35860**

Density plot for song duration

Density plot for song duration

ML ALGORITHMS : PROCESS AND EVALUATION

- PERFORMED PCA AND T-SNE COMPONENT ANALYSIS.

- PCA EXPLAINED VARIEANCE OF ONLY 32% OF DATA.

- INFERENCE : DATA HAS NON – LINEAR STRUCTUE.

- USED SCIKIT LEARN API FOR SPLITTING DATA IN 70:30

MODEL COMPARISON

Random forest

- Number of estimators :70

- Precision score : 0.761

- Recall score : 0.854

- F1 score : 0.805

- Accuracy : 0.792

- Cross-val score : 0.792

Support vector machine

- Kernel : polynomial

- Precision score : 0.640

- Recall score : 0.956

- F1 score :0.767

- Acuuracy :0.797

- Cross-val score:0.755

NEXT STEPS : FUTURE RECOMMENDATIONS

- There are many areas where we can make more deeper analysis and future work need to be performed.

- Study the features importance in each decade to see the change in trend.

- Include additional features like revenue like revenue and budget information against songs to justify the popularness of a song.

- Study influence of popularity of the artist on the popularity of a song.

- Also include the music label feature to u understand the reach and its effect on the popularity of song.

VIDEO LINK:

- https://youtu.be/h-Ctr6KVcUY