Predicting Spotify Popularity

A Tool for Industry Professionals

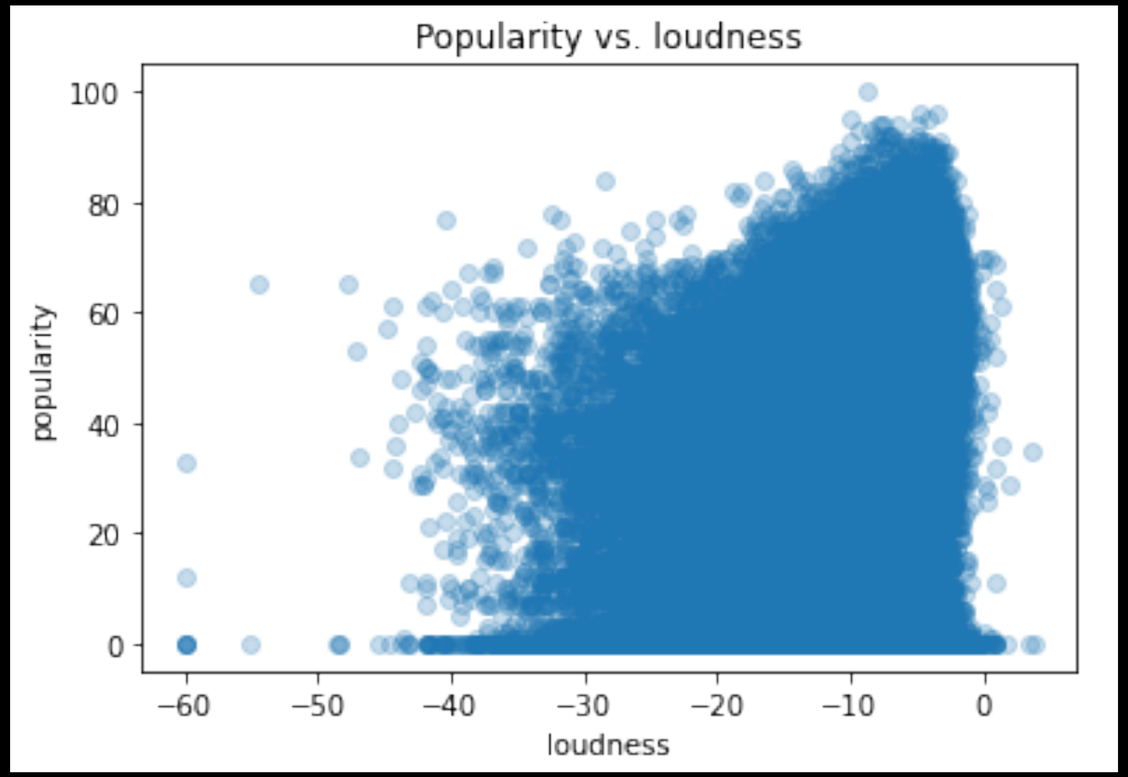
Carlos Garza

Business Problem

- In the music business today, Spotify metrics function like an artist's resume
- The pop and country markets use song selling/buying business models, but no objective song scoring methods
- A model that predicts a songs popularity can provide an edge in the industry

Data Utilized

- Data originally sourced from Spotify API
- Data gathered by Kaggle user
- Dataset contains ~175K songs
- EDA sheds light on “loudness wars”



Baseline Models

- Various machine learning models auditioned with default parameters to compare baseline performance
- Models evaluated by r2 score and two similar measurements of error

Baseline Test Performance	R2 Score	RMSE	MAE
Random Forest	0.589	12.178	8.943
XGBoost	0.575	12.382	9.106
Neural Network	0.519	13.175	9.390
KNN	0.415	14.529	10.811
Decision Tree	0.162	17.382	12.025

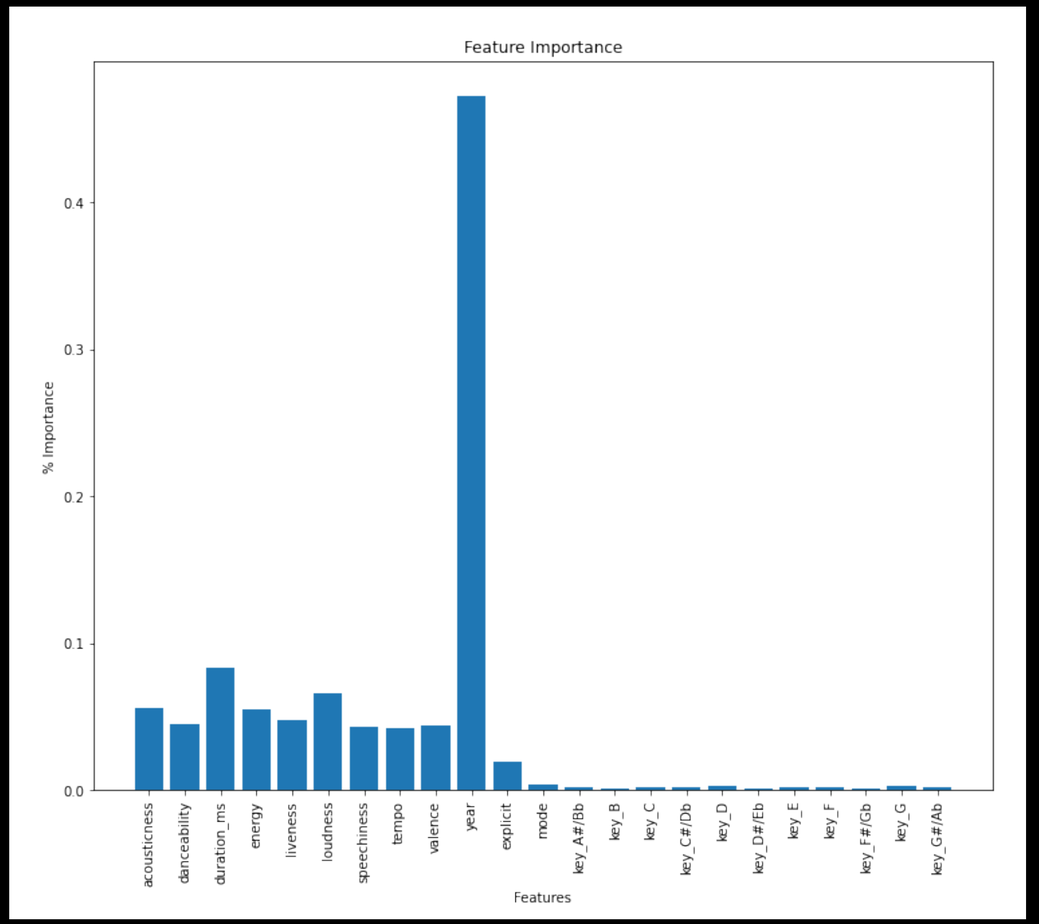
Model Selection

- Random forest outperformed other baseline models.
- Model tuned to maximize performance

Random Forest	R2 Score	RMSE	MAE
Train Data	0.902	5.936	8.943
Test Data	0.589	12.173	8.822

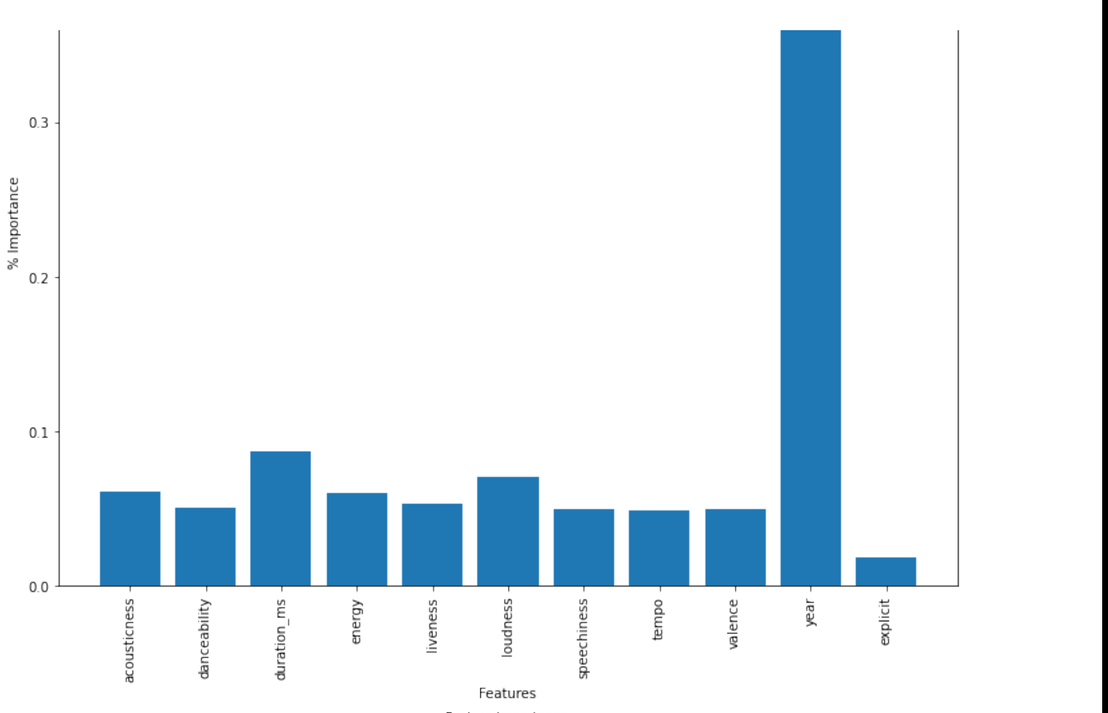
Results

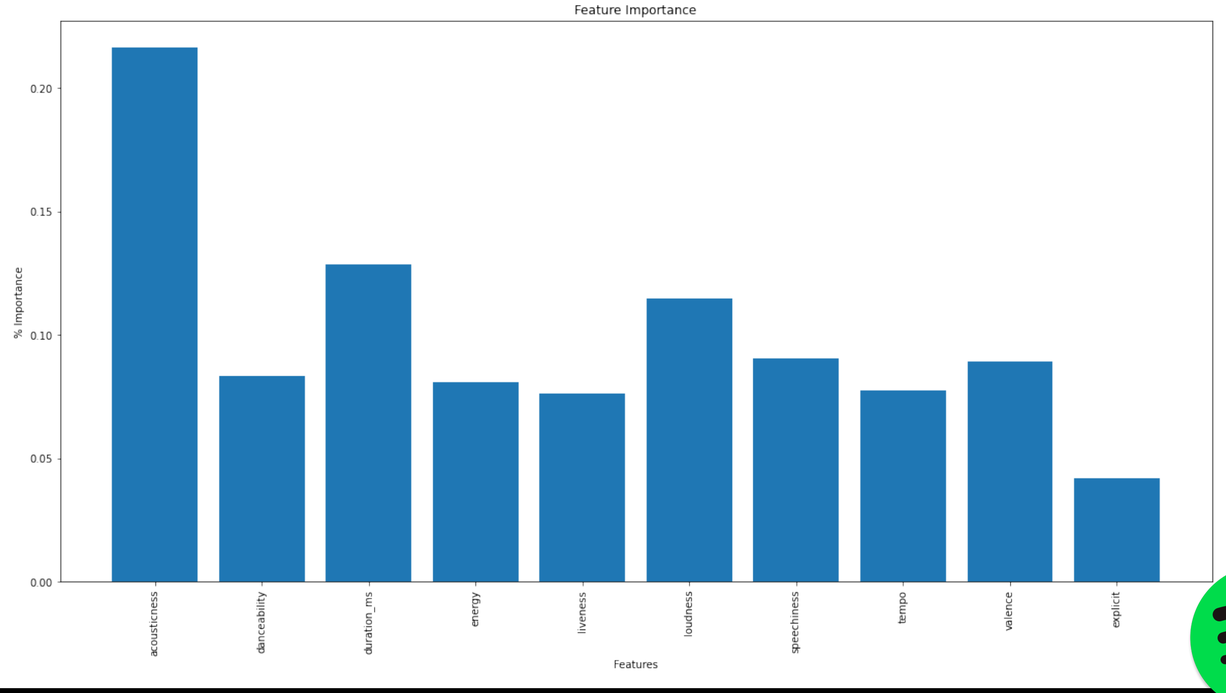
- Model explains 58% of the variance of the data
- Year released is most important feature
- Song key is least important feature



Results

- Feature importance validated by observing a similar model that omits song key and another similar model that omits key and year released.





Conclusions

- Model offers valuable metric to an otherwise subjective part of the music industry
- Music release year is the most influential predictor of success on Spotify platform. Emulations of different years of music may be a useful strategy
- Song key does not matter. Western music uses scales with even temperament no conceivable difference exists between keys to the average listener

Future Work

- Refine model to take genre into consideration
- Explore how Spotify quantifies continuous attributes (energy, danceability, etc)
- Revisit neural network models
- Develop model that can take .wav or .mp3 input

Thank You

carlosgarza.io