
Search
Home
Library
Spotify Capstone
Utilizing Spotify API:
Exploring Popularity and Genres
Based on Audio Features
Samuel Hewitt

Table of contents

Spotify Capstone

Search

Home

Library

01

Introduction

Background and Problem Statement

Spotify Capstone

Search

Home

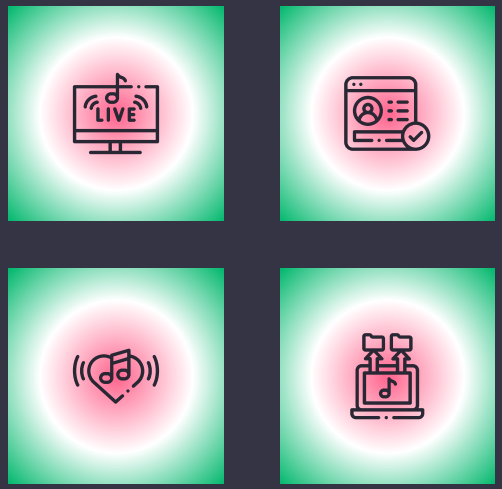
Library

Spotify Web API

Follow

Data collected using Spotify's Web API in hand with Spotipy, a lightweight Python library for Spotify Web API. Both require client credentials to collect data. Song data from past 10 years have been collected

Problem Statement



Music streaming services have changed the metrics for a song/artist to be considered successful. The goal of this project is to determine if one can predict a song's popularity and success. I also aim to determine if a song's genre can be predicted based on it's audio features with disregard to the artist who made it.

Track Audio Features

Order by

#	Feature	Description	Range	Average
1	Popularity	Popularity of track	(0.0, 1.0)	46.49
2	Danceability	How suitable track is for dancing	(0.0, 1.0)	0.60
3	Energy	Measure of intensity and activity	(0.0, 1.0)	0.66
4	Key	The key track is in	(-1, 11)	5.38
5	Loudness	Overall track loudness in decibels (dB)	(0, 1)	-7.59 (dB)
6	Mode	Modality of track Major = 1, Minor = 0	(0, 1)	0.63
7	Speechiness	Confidence of spoken words present	(0.1, 1.0)	0.098

Track Audio Features

Order by

#	Feature	Description	Range	Average
8	Acousticness	Confidence track is acoustic	(0.0, 1.0)	0.22
9	Instrumentalness	Predicts if tracks contains no vocals	(0.0, 1.0)	0.12
10	Liveness	Detects presence of audience	(0.0, 1.0)	0.22
11	Valence	Musical positiveness conveyed	(0.0, 1 .0)	0.49
12	Tempo	Estimated tempo in BPM	(-1, 11)	123.06
13	Duration	Track duration	(0, ∞)	212.30 (s)
14	Time Signature	Amount of beats per bar (3/4, 7/4)	(3, 7)	3.93

Spotify Capstone

Search

Home

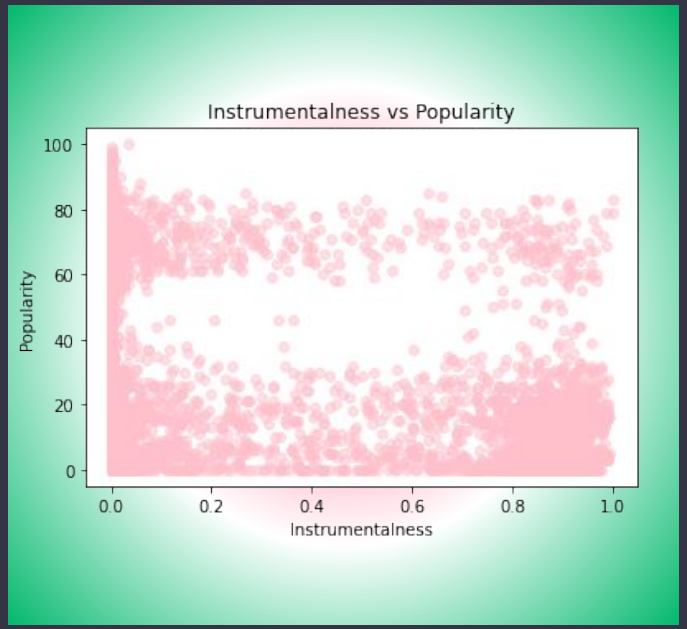
Library

02

Exploratory Data Analysis

EDA

Order by



- Instrumentalness has largest correlation with popularity
  - -0.37 correlation
  - Clear gap in middle and upper portion
  - Most popular songs are not instrumental
- Largest positive correlations
  - Loudness: 0.25
  - Danceability: 0.19
- Remaining audio features have correlations close to 0

Spotify Capstone

Search

Home

Library

03

Modeling

Regression, Clustering, Classification

Modeling - Popularity by Audio Features

Order by

#	Model	Train R2	Test R2	Testing MSE	Baseline MSE
1	LinearRegression	0.2227	0.2103	754.75	955.76
2	KNeighbors	0.3731	0.0605	897.96	955.76
3	DecisionTree	0.9961	-0.2380	1183.24	955.76
4	Bagging	0.8710	0.3007	668.38	955.76
5	RandomForest	0.9069	0.3497	621.52	955.76
6	AdaBoost	0.2226	0.2080	756.96	955.76

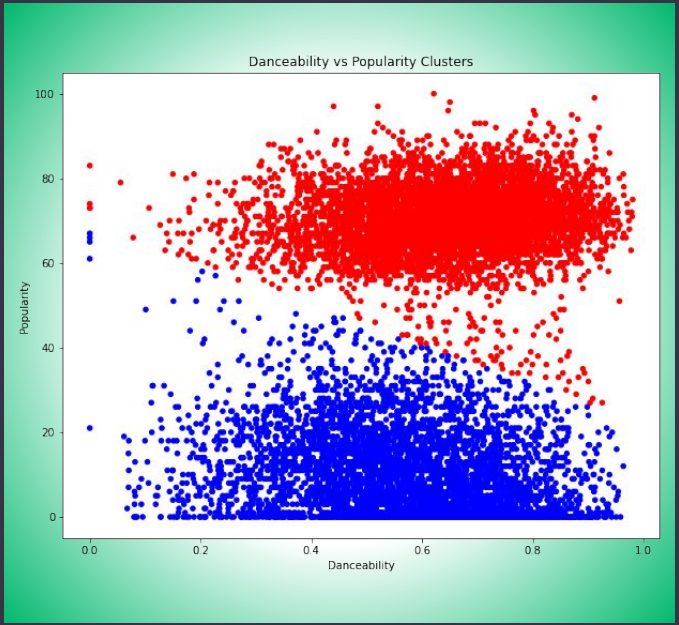
Modeling - Popularity by Audio Features & Genre

Order by

#	Model	Train R2	Test R2	Testing MSE	Baseline MSE
1	LinearRegression	0.8483	0.6401	343.98	955.76
2	KNeighbors	0.3047	0.0605	897.96	955.76
3	DecisionTree	0.9961	0.0619	896.56	955.76
4	Bagging	0.9031	0.4511	524.66	955.76
5	RandomForest	0.9295	0.5087	469.54	955.76
6	AdaBoost	0.1827	0.1788	784.90	955.76
7	RidgeCV	0.8124	0.7015	285.31	955.76

Classifying Popularity

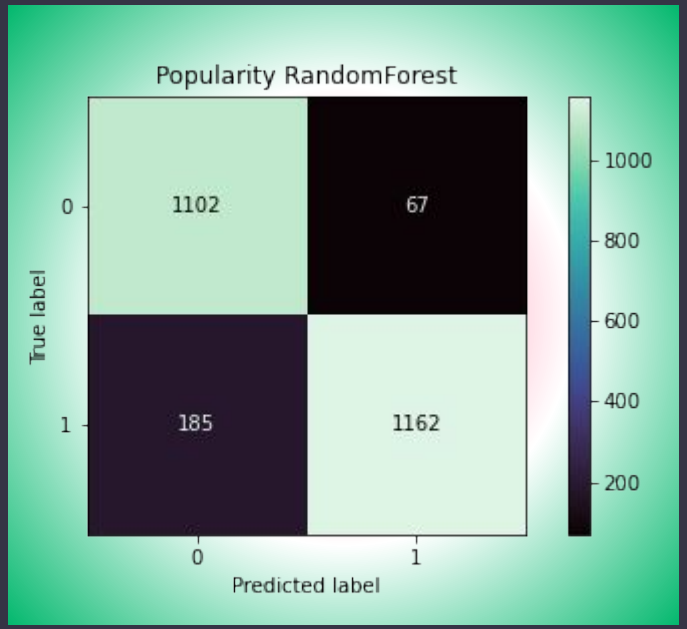
Order by



- KMeans provided the best clustering model
  - Datapoints too clustered to for DBScan with low epsilons
- Silhouette Score: 0.4905
- Clear separation between popular and not so popular songs
  - Average popularity score: ~46.45

Classifying Popularity

Order by



Model	Training Score	Testing Score
LogReg	0.5374	0.5378
RandomForest	0.9980	0.8898
AdaBoost	0.8490	0.8335
GradientBoost	0.8503	0.8446

Classifying Genres

Order by

Classification Model	Training Score	Testing Score
LogReg	0.0824	0.0758
RandomForest	0.3416	0.0225
AdaBoost	0.0723	0.0675

- ~1,300 genres in the dataset
  - Most prominent genres: Pop and rap
  - ~450 single instances
    - ie Virgin Islands Reggae, German Stoner Rock, Classic Greek Pop

Spotify Capstone

Search

Home

Library

3.5

Tableau Dashboard

Link will be provided, feel free to play around and find some new music!

Spotify Capstone

Search

Home

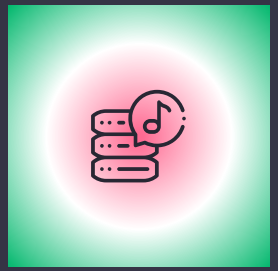
Library

04

Conclusion

Recommendations and Next Steps

Conclusion



Predicting Popularity

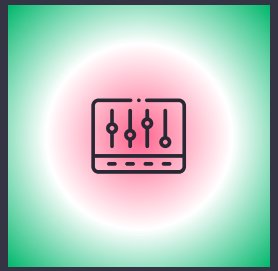
Audio features cannot accurately predict popularity. Difficult to predict a genre based on audio features.

Outside Factors

Other media's influences on trends (ie TikTok, Netflix, etc.)

Artist's behaviors can influence popularity (ie reemergence, death, controversy)

Recommendations



Collaboration

Collaborating with those who can influence a song's popularity, whether that be the artist directly or media powerhouses

Trends

Observe trends (ie TikTok, popular shows, artist behavior) to predict potential popularity

Search

Home

Library

Spotify Capstone

Thanks!

Feel free to ask any questions