



SoundCloud Project

Predict the popularity of new songs

Agenda

- **Introduction**
- **Data Collection & Preprocessing**
- **Data Exploration**
- **Model Selection**
- **Result**

SoundCloud



- ▶ Intended to allow musicians to collaborate
- ▶ But transformed into **a publishing tool** for music distribution
- ▶ In 2014, started to **monetize their content** through pre-roll audio ads, mobile display ads, and native content
- ▶ Key features: access uploaded files via **unique URLs**, thus allowing sound files to be embedded in Twitter and **Facebook** posts



Home

Stream

Collection

Search



Upgrade

Data Collection

Top 50

USA

All music genres

Top 50

All music genres

New & hot

All audio genres

#

Track

1

Lil Baby
Drip Too Hard

MUSIC

2

Alternative Rock

Ambient

30 genres X 50 songs = 1500**Top 50s in US**

3

Deep House



LIL YACHTY

4

Disco

Flipp Dinero
Leave Me Alone (Prod. by Young Forever x Cast Beats)

Drum & Bass

5

Dubstep

Kodak Black
Calling My Spirit

Electronic

Folk & Singer-Songwriter



16yrold



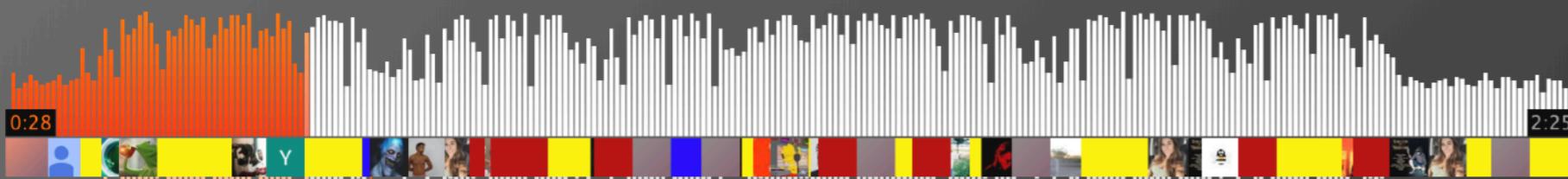
Lil Baby

Drip Too Hard

4 months ago

Hip Hop

PARENTAL
ADVISORY
EXPLICIT CONTENT



...BiG...C... check my shit out, appreciate the feed back on comments. i will also rat...



Lil Baby

80.6M 1.09M 31.5K

Release date:
12 September 2018

18,335 comments

325K 83

Lil Baby



Drip Too Hard

4 months ago

Hip Hop

PARENTAL
ADVISORY
EXPLICIT CONTENT

Input Variables

of likes
of reposts
of comments
of followers
of tracks
length of song
release date

Output Variables

of plays
(popularity)

Lil Baby

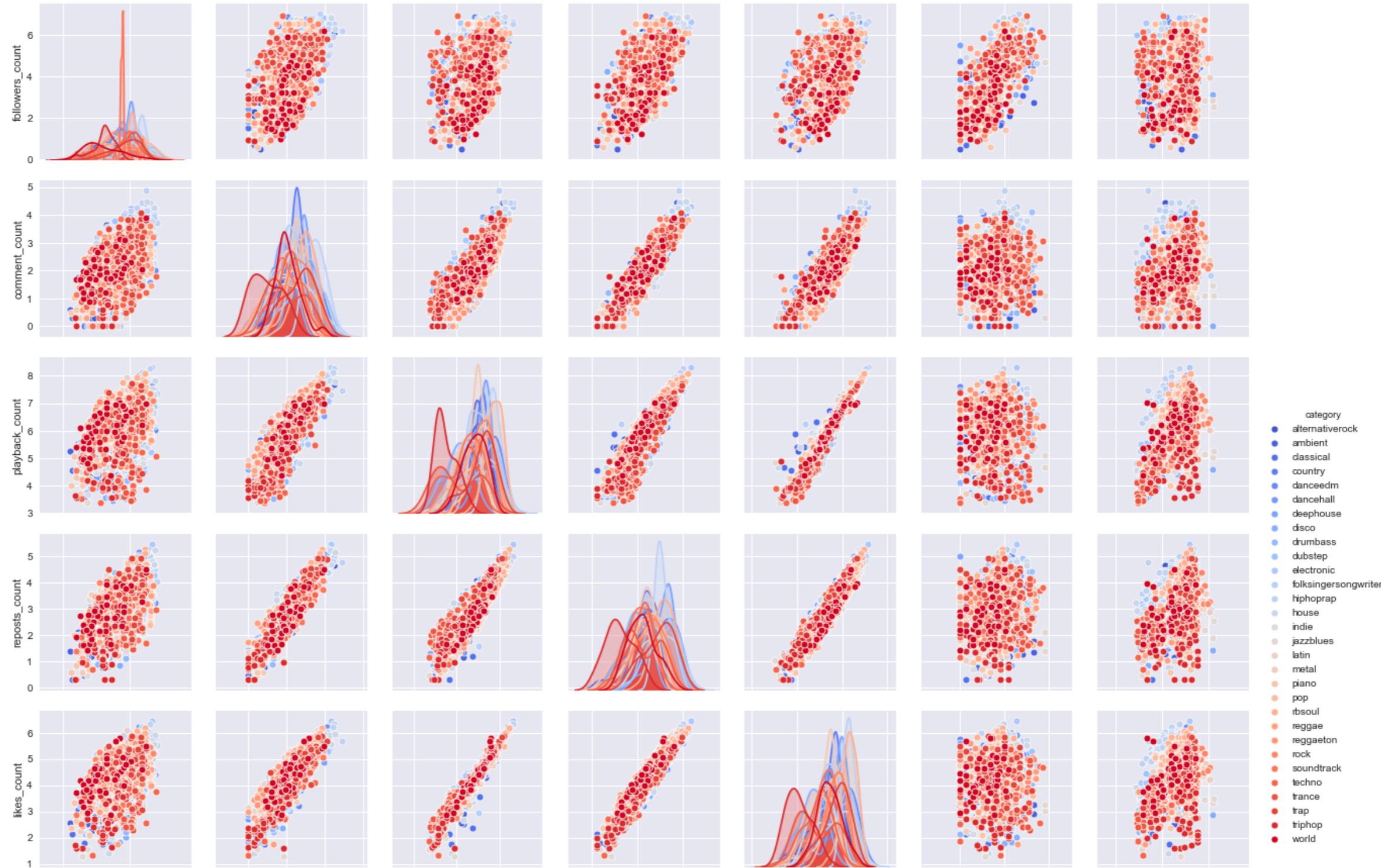
10,555 comments



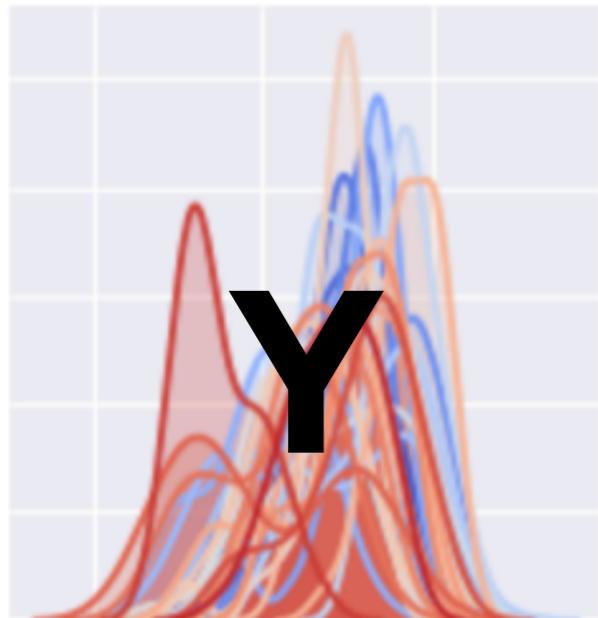
325K

83

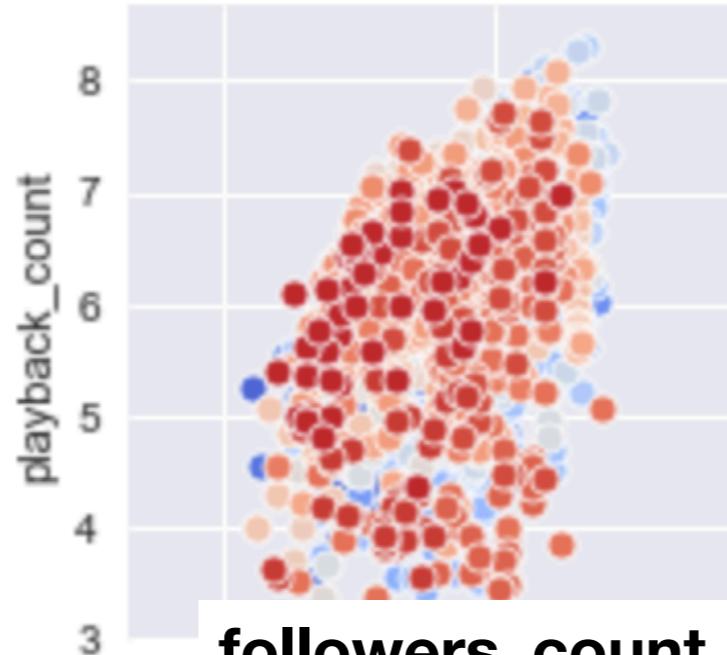
WOW



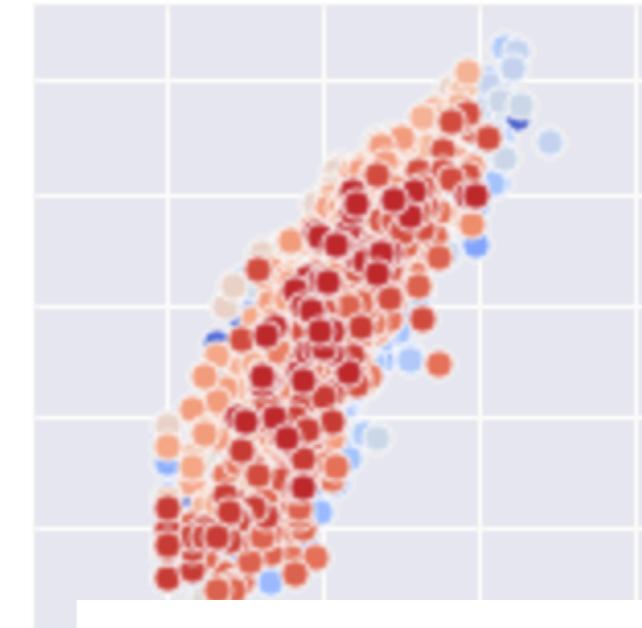
WOW 😱



playback_count



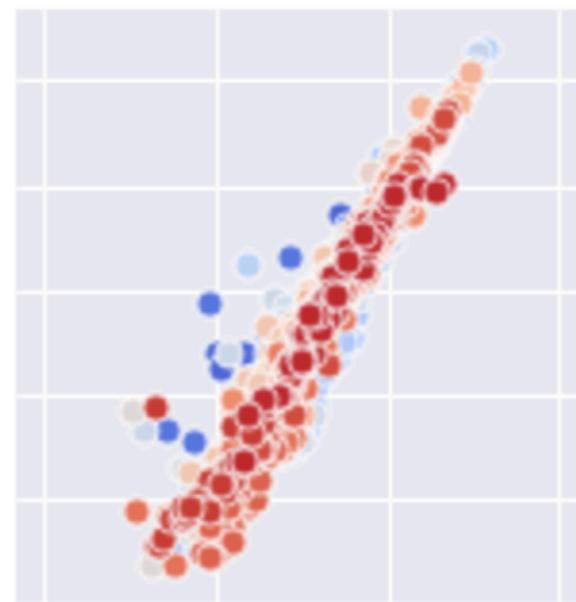
followers_count



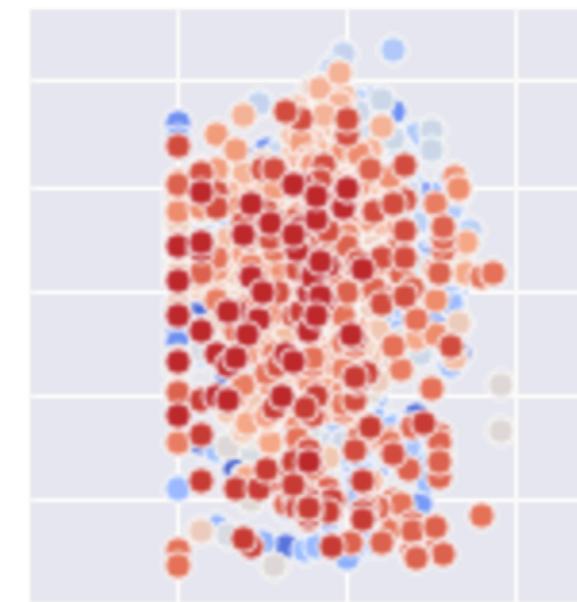
comments_count



reposts_count



likes_count

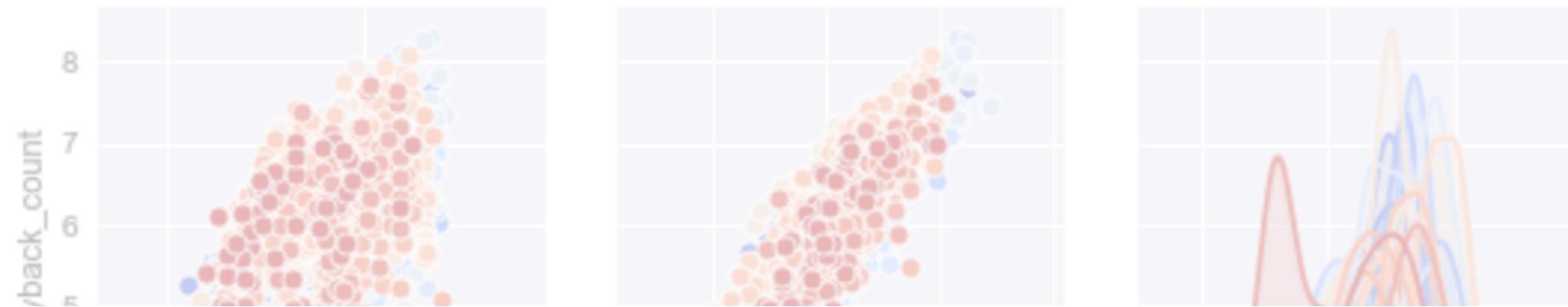


tracks_count



age

Too good to be true



Linear Regression val $R^2: 0.974$

Ridge Regression val $R^2: 0.974$

2 polynomial regression val $R^2: 0.976$



Trippie Redd ★

lil-14 the Savage

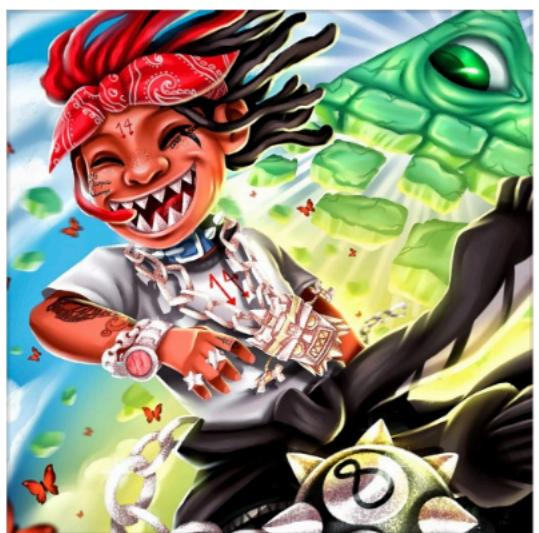
1400 1400 1400

All Tracks

Albums

Playlists

Reposts



Trippie Redd

A Love Letter To You 3 Album · 2018

2 months ago

Hip-hop & Rap

3:35



1 Topanga

► 19.2M



2 Fire Starter ft. Emani22

► 2.89M

All

Tracks

Albums

Playlists

Reposts



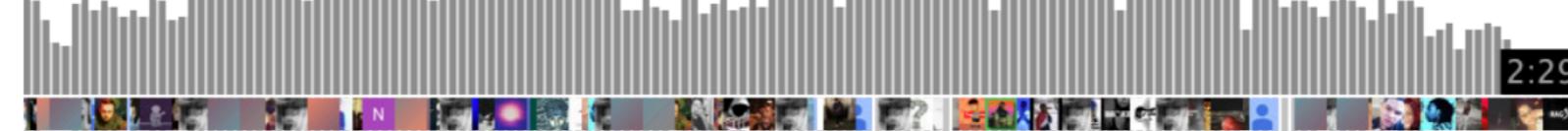
Trippie Redd

Camp Fire Tale

2 months ago

Hip-hop & Rap

2:29





1500 songs -> 600 songs

Input Variables

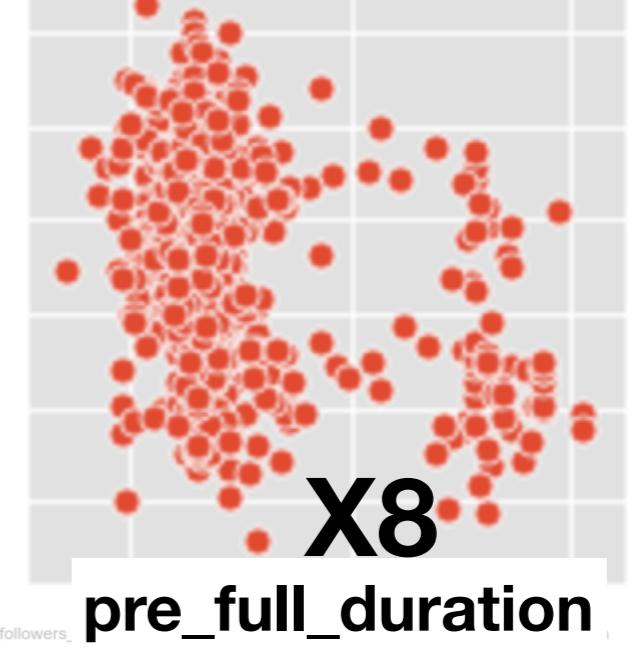
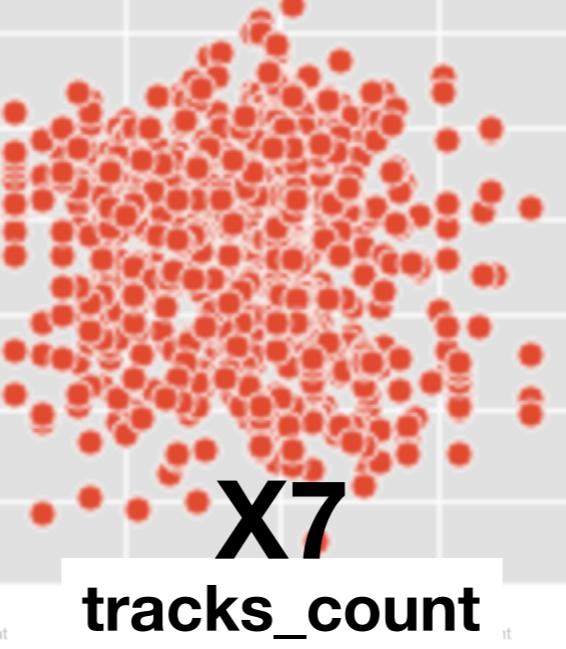
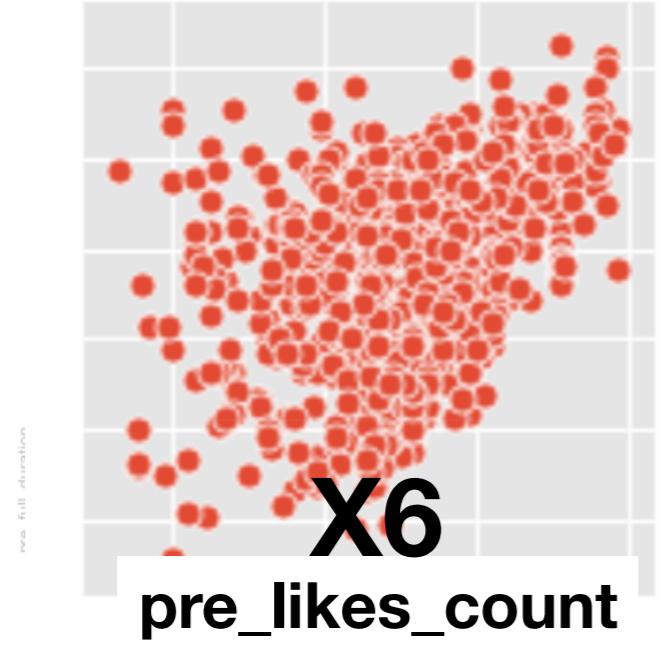
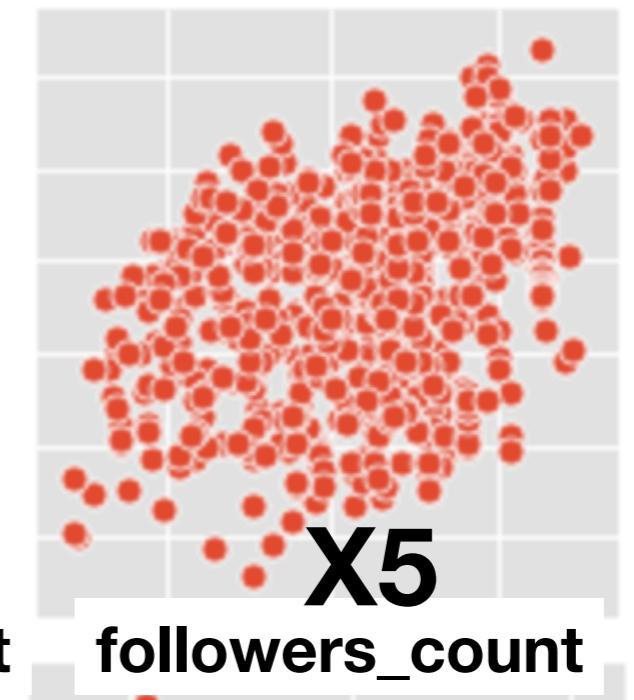
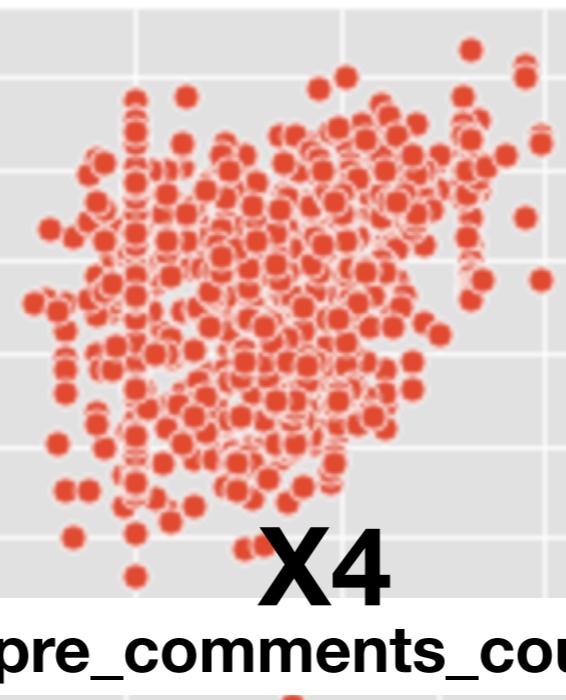
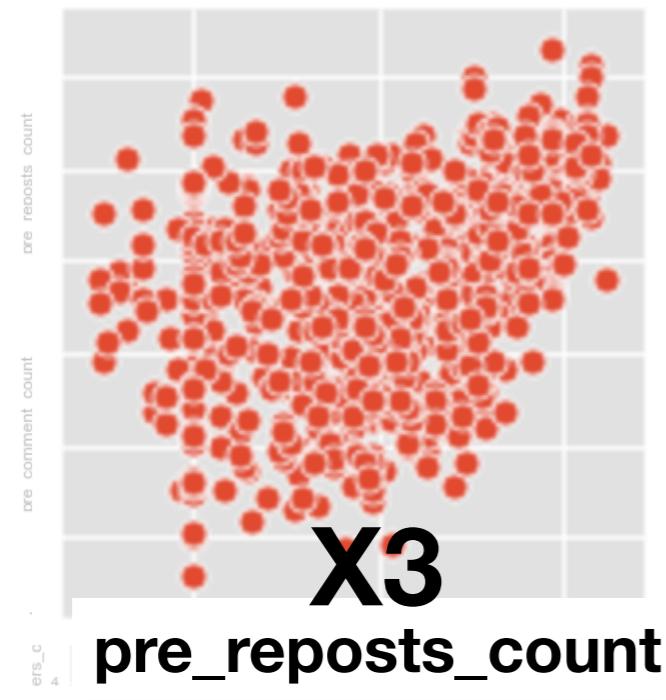
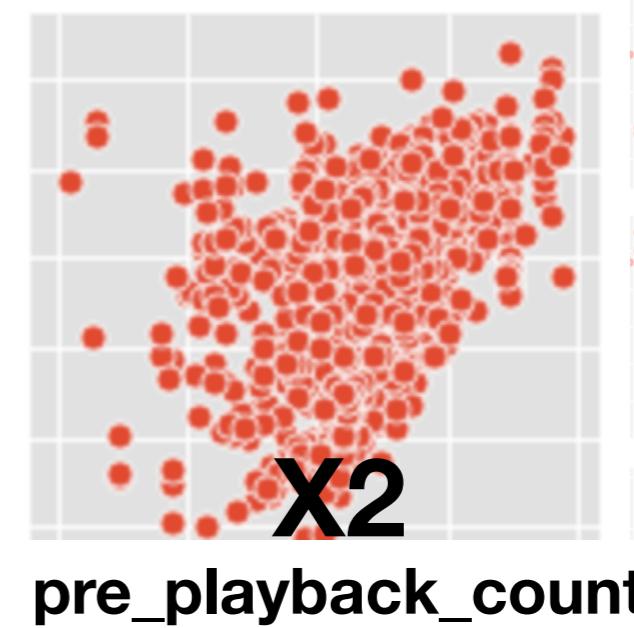
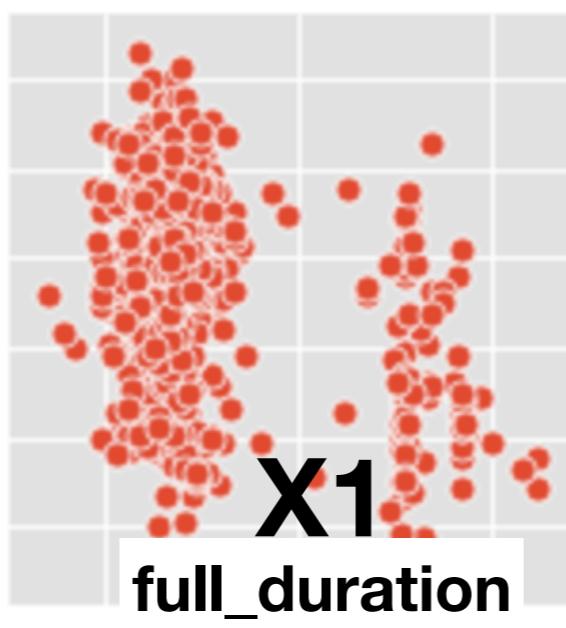
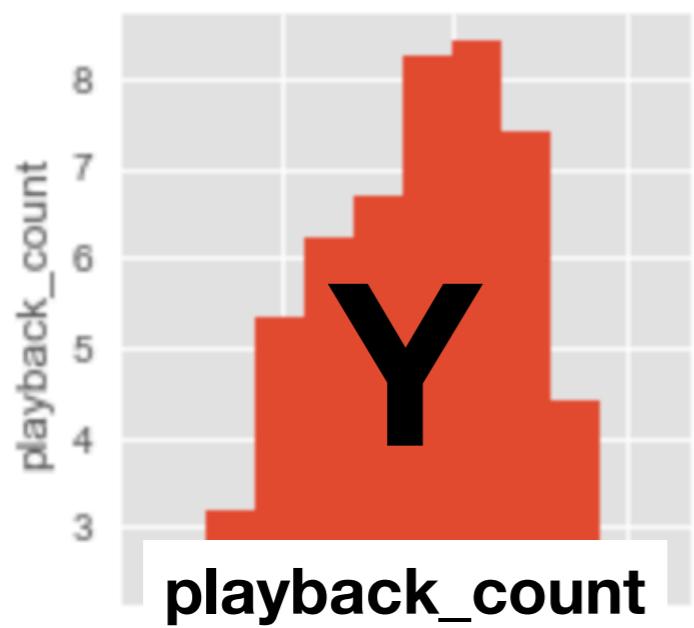
- # of pre_plays
- # of pre_likes
- # of pre_reposts
- # of pre_comments
- # of tracks
- # of followers
- Length of a song

Output Variables

of plays

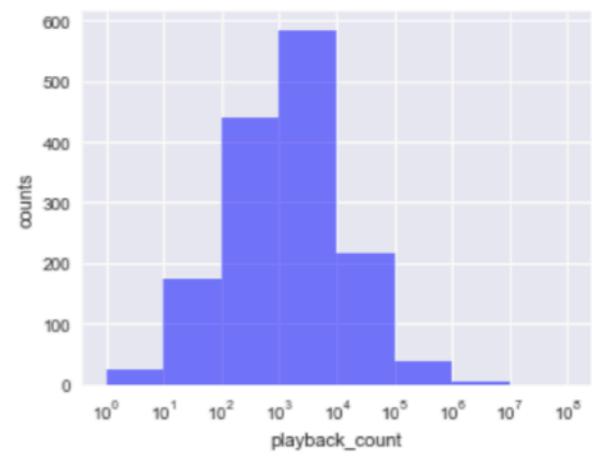
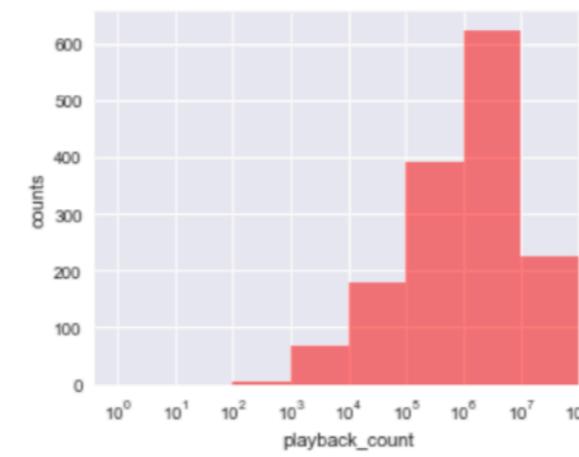
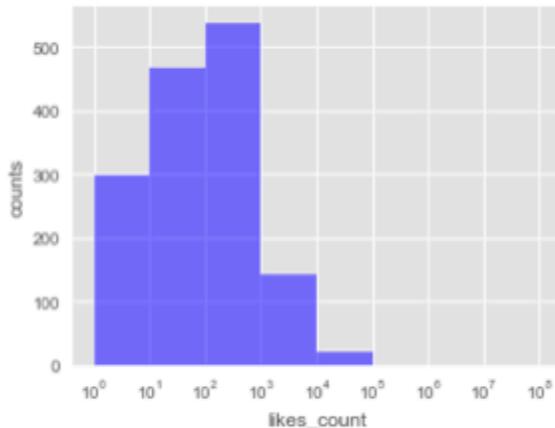
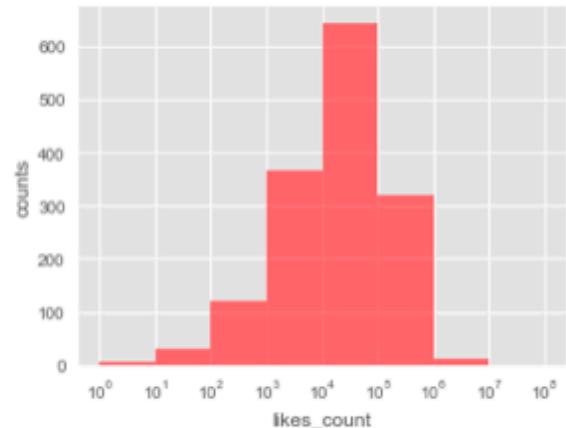
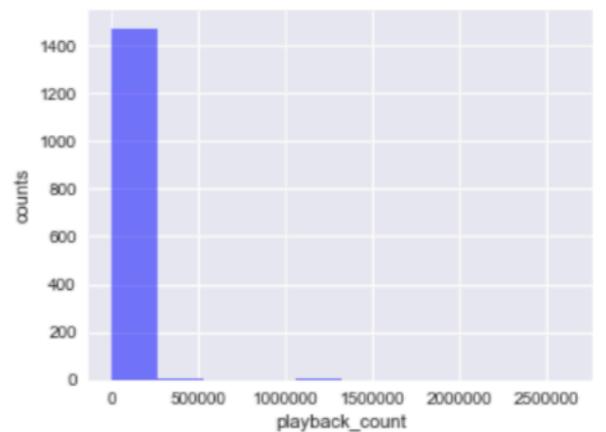
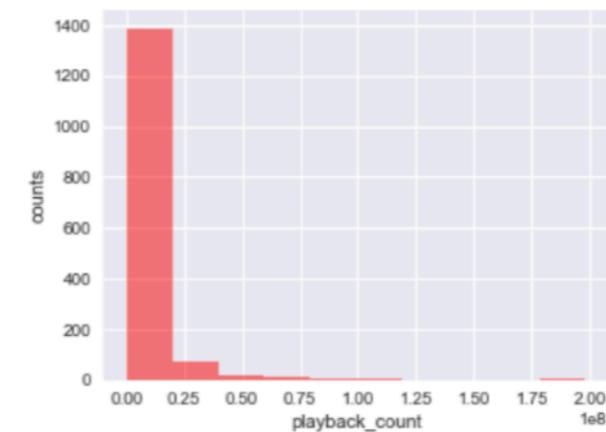
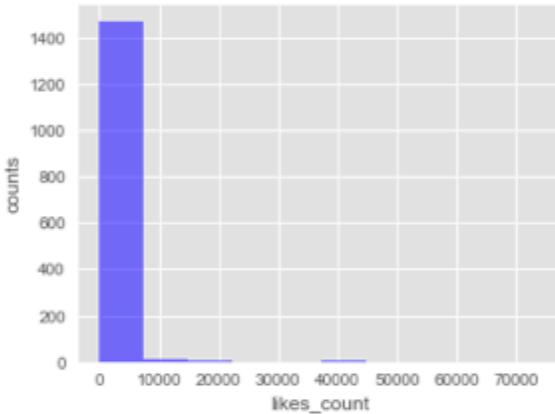
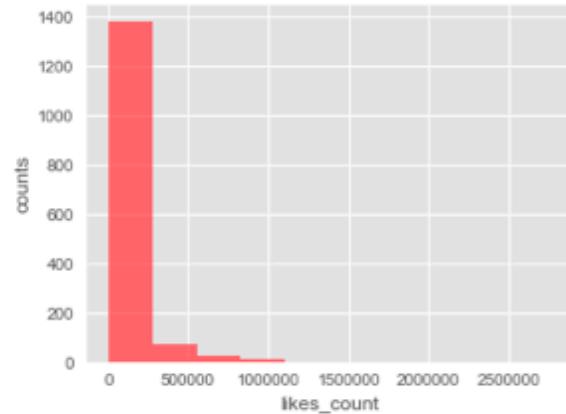
Lil Bab

*the data we can get
before a song is released*



Methodology

☁ Log transformation for input & output



Methodology

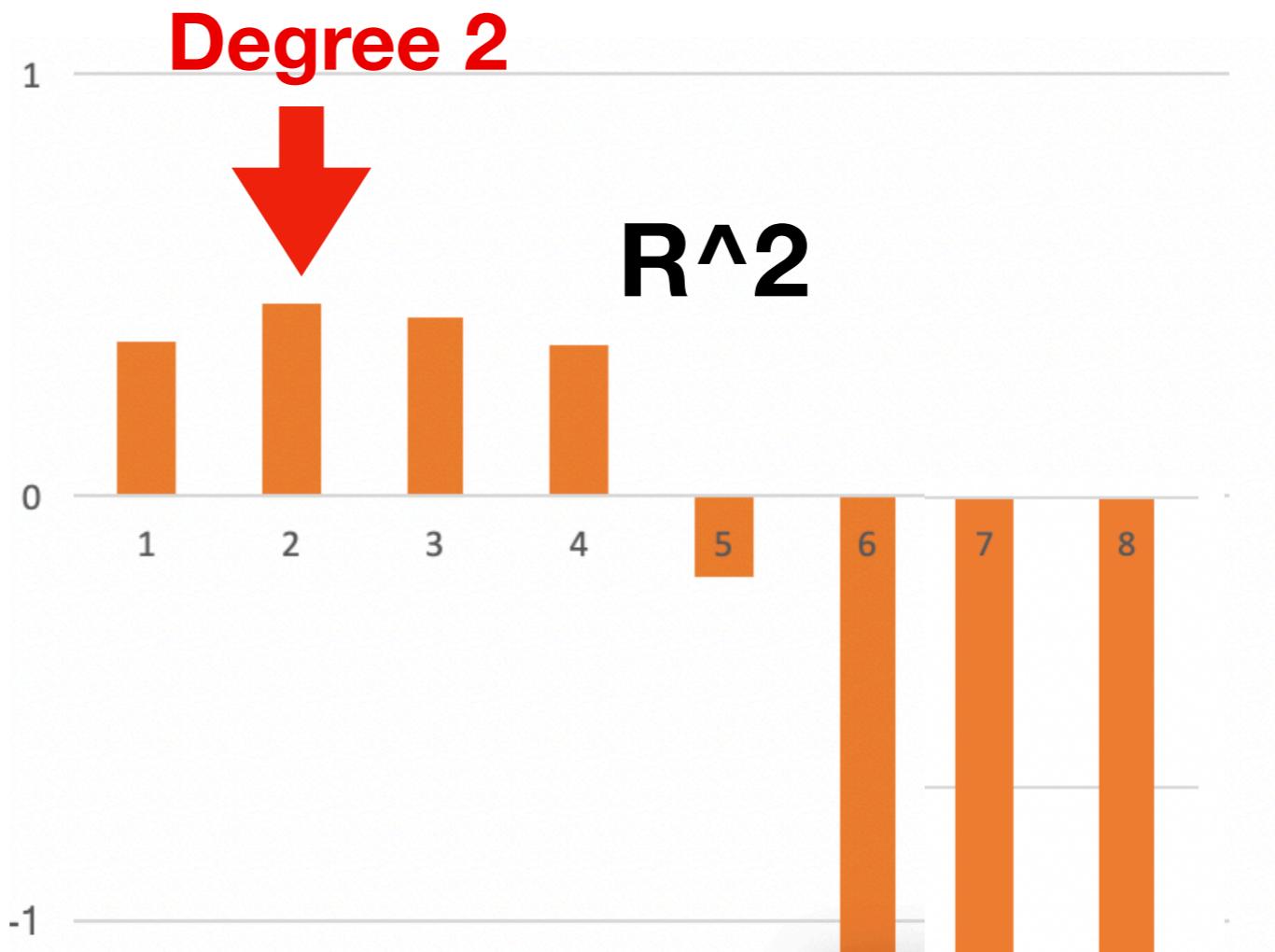
- Log transformation for input & output
- -inf -> 0
- .20 hold out
- .70 test .30 training

Methodology

degree	R^2_test
1	0.371
2	0.482
3	0.421
4	0.36
5	-0.191
6	-5
7

心脏病 R^2 holdout

0.45 ~ 0.48



Results

High coef:

pre[▶]plays, followers



R^2:

0.45



of plays / popularity

\$\$
Ads

\$\$
Streaming cost

Release date

Future research

More data with better scaling

Audio analysis features: tempo, loudness, key

Artist related features: artist location

Song related features: released date-seasonality, genres, lyrics



Thanks