




# ARTIST RECOMMENDER SYSTEM








JON ECKSTEIN  
GA-DAT-13


# PROBLEM


- Dropped echonest, lost ability to get similar artists
- Every major player has this feature
- Get it back using our data



RELATED ARTISTS







-  Amon Tobin
-  Nightmares On Wax
-  The Herbaliser
-  Jon Kennedy
-  Blockhead
-  Mr. Scruff
-  The Cinematic Orchestra




 Bonobo

Albums Songs Biography [Related Artists](#)


Related Artists

-  Funk Porcini  
15 albums, 140 songs
-  Coldcut  
12 albums, 107 songs
-  Jon Kennedy  
27 albums, 218 songs
-  Nostalgia 77  
23 albums, 189 songs
-  Skalpel  
7 albums, 67 songs
-  Kruder & Dorfmeister  
5 albums, 49 songs




Tracks Albums **Similar Artists** Listeners Po


[Contemporaries](#) Influences Related




Cinematic Orchestra




Grasscut




Bibio



Koushik



Nostalgia 77



Quantic

# DATA SET/METHODOLOGY

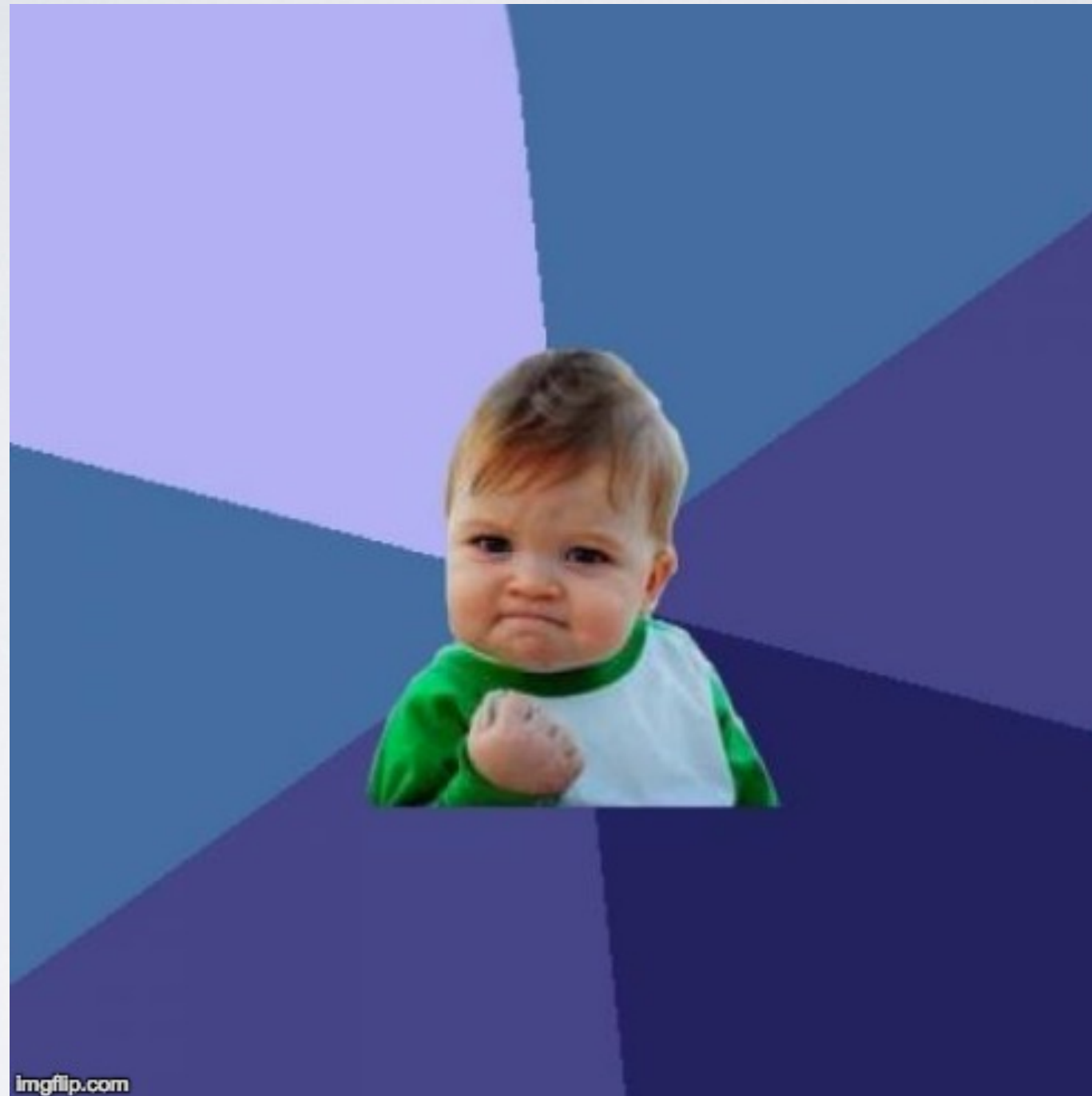
- Company data-set of user listens

user\_id | artist\_id | listen\_count

- iTunes API - Genres
- python-recsys library
  - collaborative filtering approach



# MY PLAN



# REALITY



# WHAT WORKED:

# CHALLENGES

- Lack of data
- Long tail
- Validation



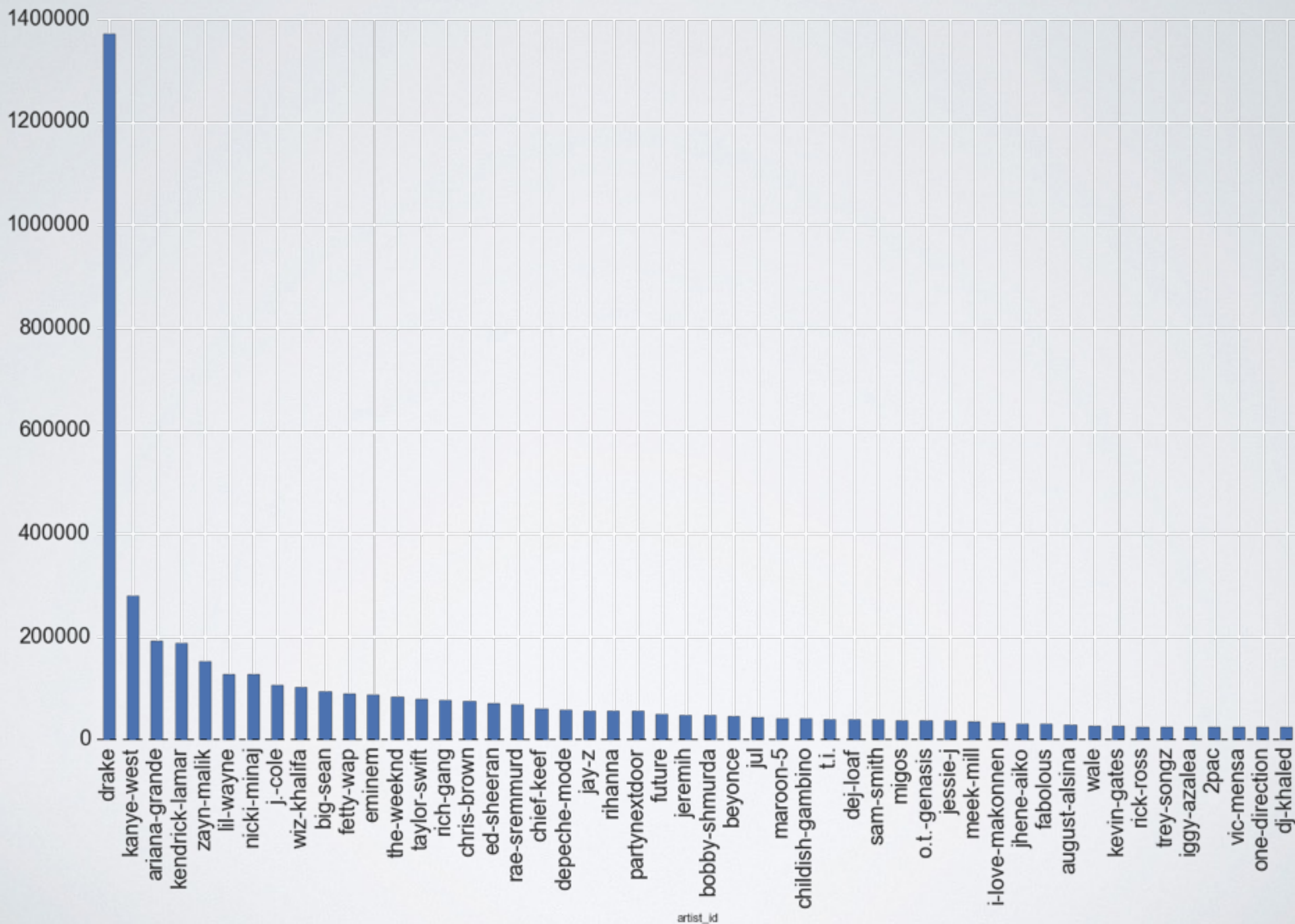
# LACK OF DATA

- Ultimately the biggest problem
- Matrix too sparse
  - Density = 0.0005%,
  - Movielens density = 4.5%
- Attempt at solving this was trying to trick the algorithm

# LONG TAIL

- 90% of listens belong to 4% of artists
- The majority of the artists don't have enough data for good recommendations

Listen counts for top 50 artists



# LONG TAIL CONTINUED

## **Drake:**

DJ Kahled

Lil' Wayne

2 Chainz

DJ Mustard

J. Cole

Big Sean

## **Sturgill Simpson:**

The Cadillac Three

Hans Zimmer

James Horner


Naughty By Nature

The Naked and Famous

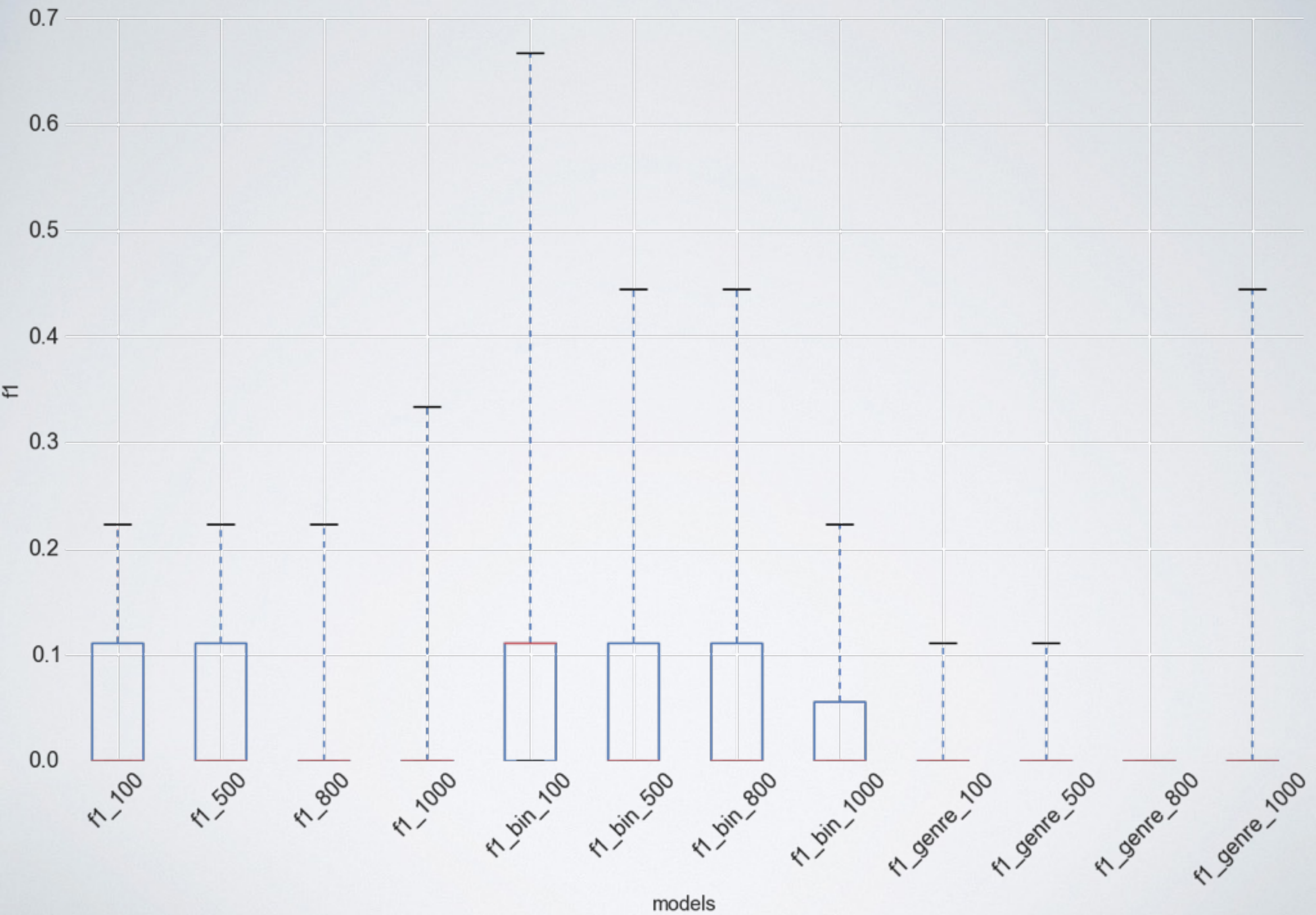
Erasure



# VALIDATION

- How to validate if no ground truth?
- Using Last.fm API The Last.fm logo, which consists of a red circle containing the white letters 'as' in a stylized font.
- Also using collaborative filtering via scrobbling
- Last.fm recommendations for top 200 artists, normalized names via “slugify”
- python-recsys fl score

# F1 Scores for model approaches



# WHAT'S NEXT

- Get more data
- Continually check back on my accuracy.