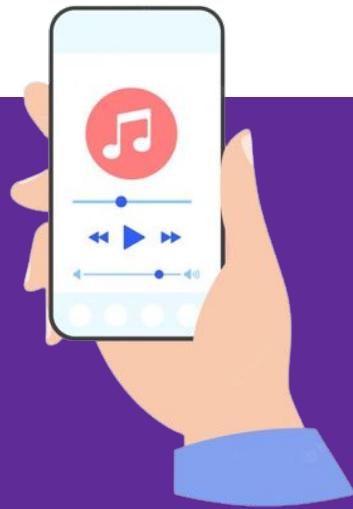


# Creating and Deploying a Music Recommender

Mitch Allison  
10/6/22



# Overview



- Business Problem
- Data
- Model Selection
- Deployment Demo
- Next Steps

---

# Business Problem



- 84% of industry revenue
- 82.1M US subscribers
- Stream 75 min/day
  - Cater to individuals

# Data

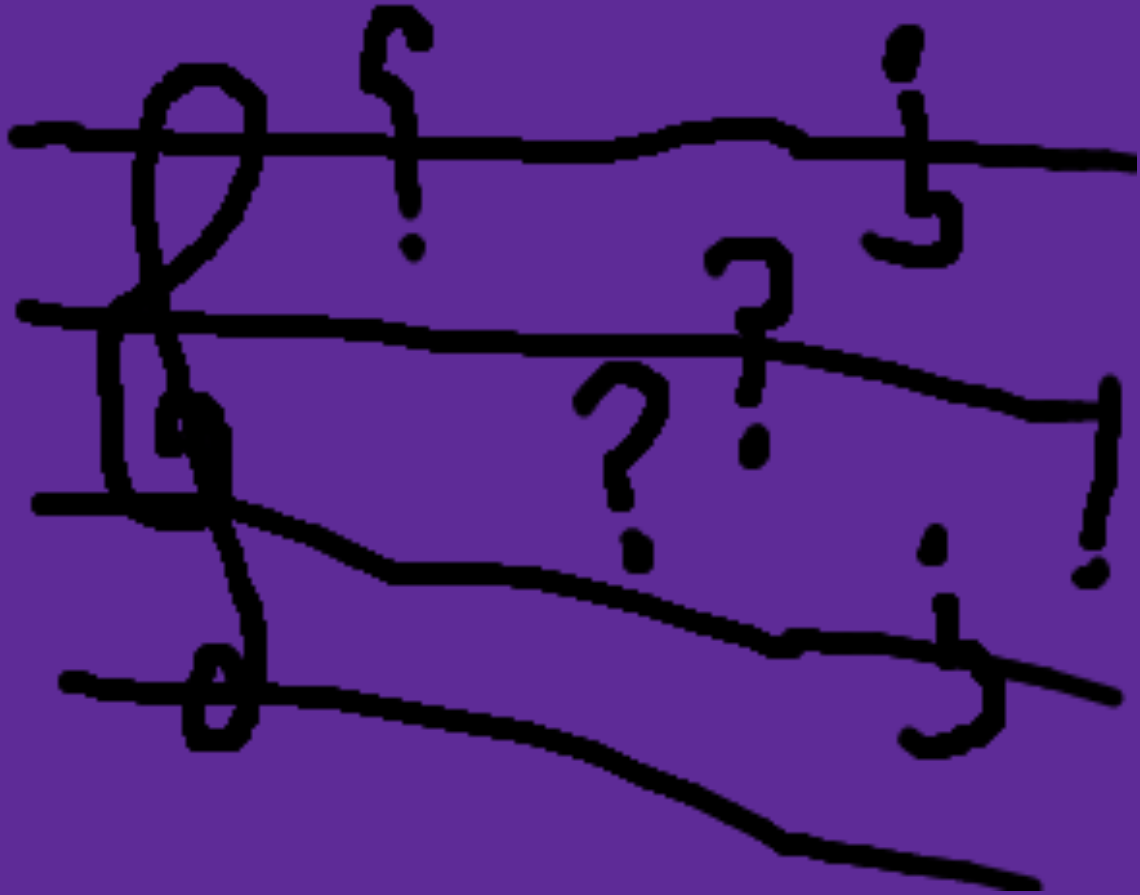


**ListenBrainz**

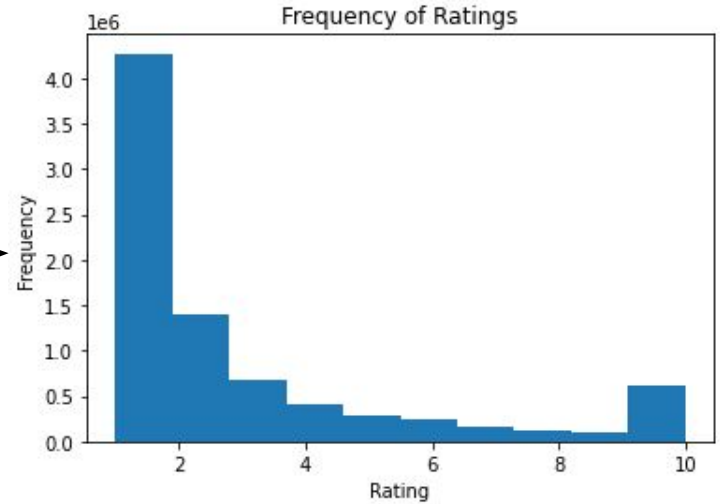
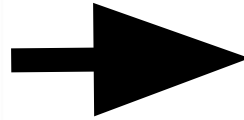
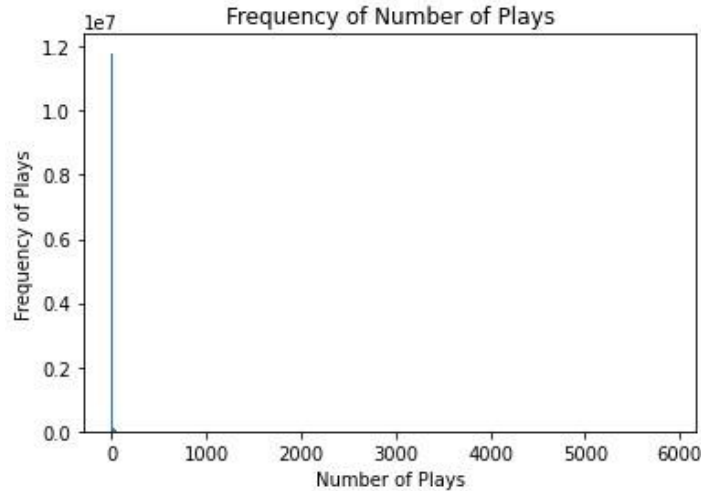
- Open-source
- 27 Gb (149M rows)
  - 37M rows used
- From 2005-2018
  - 2016-2018 used
- Unrated plays

---

How to  
create  
explicit  
ratings?



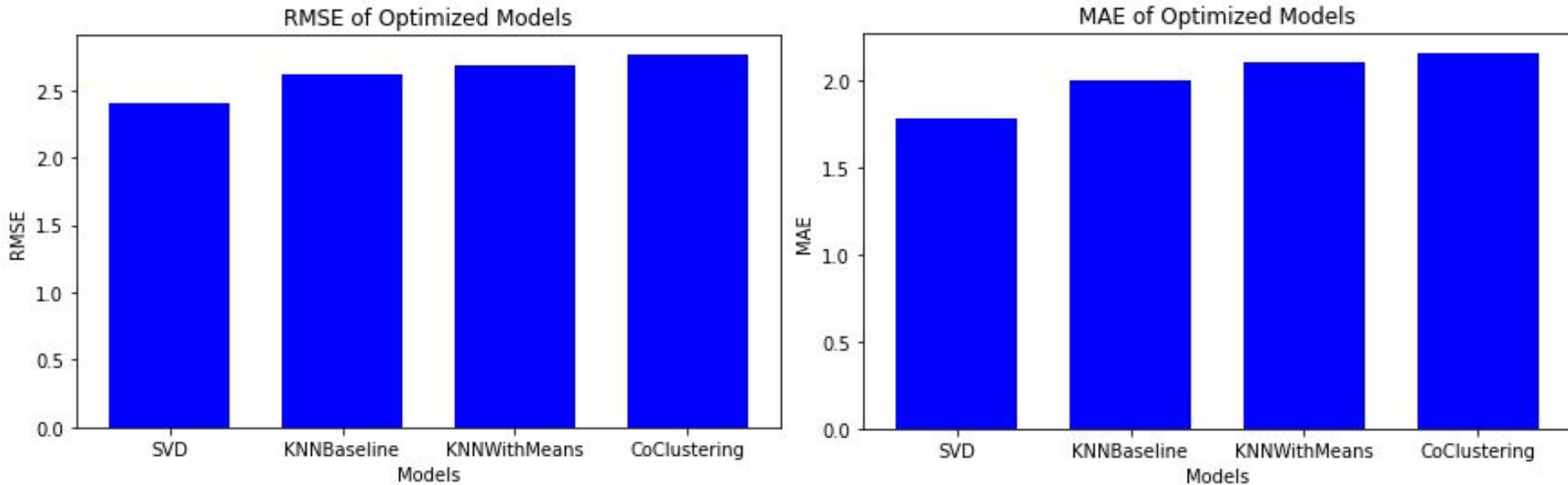
# Data Analysis / Rating Engineering



- A person listens to a song 1 time: **51%**
- Recommend a song you listened to once?

- Rating = # of plays per profile, capped at 10
- Best performing rating system

# Model Selection



- Focus on collaborative filtering
- Grid search CV optimized

# MODEL TOO BIG

- SVD model too large: 1.3 GB
  - Top 1M songs, 2.3K users
    - 8M ratings: 29M listens

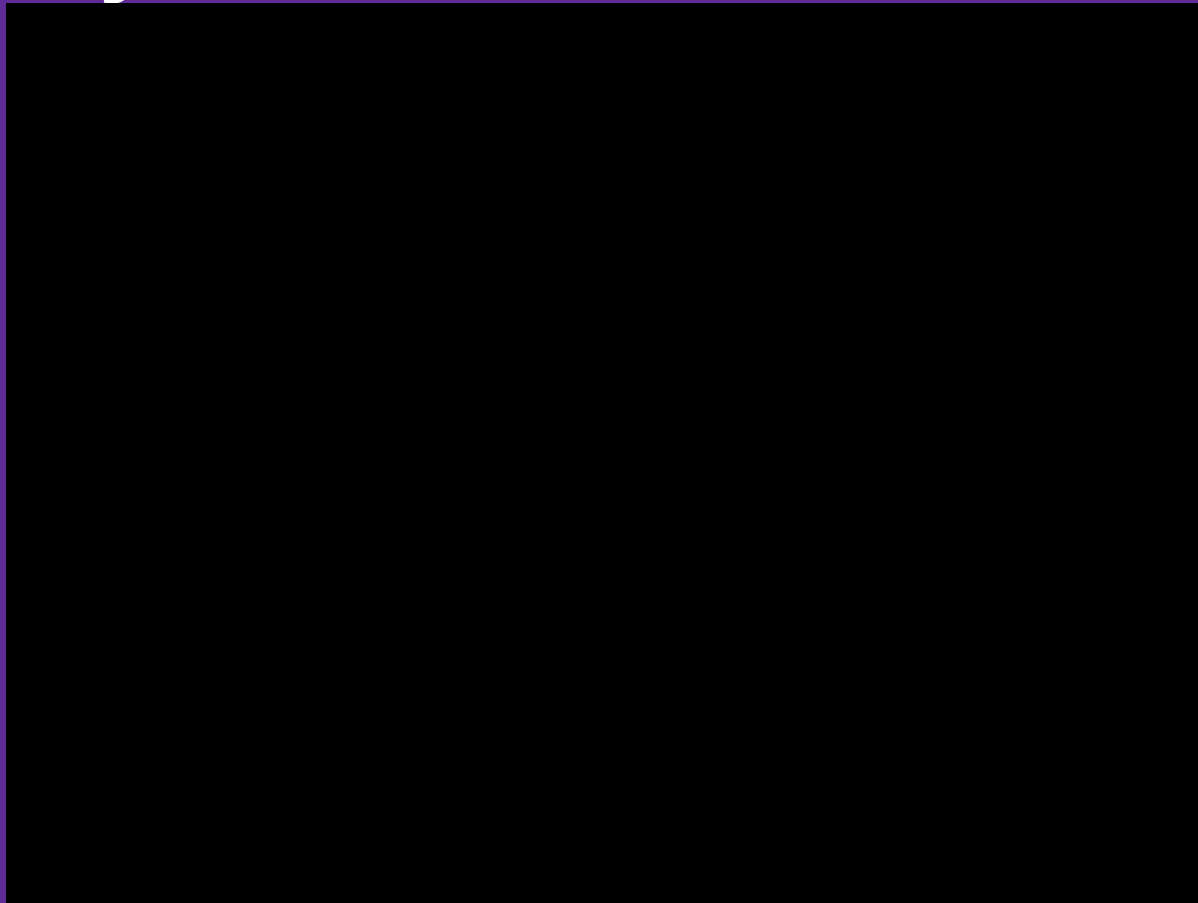


# Deployable Model



- Model: SVD
- Data:
  - Top 10k songs
  - 1.2M ratings
  - 2.3K users
- Final RMSE: 2.42
- Deployable: 41Mb

# Deployment Demo: MitchMusic



# Next Steps



- Improve model
- Train model on full dataset
- Migrate to distributed implementation
- Add additional features

---

# Questions?



Mitch Allison

[Github](#)

[LinkedIn](#)

[Email:](#)

mitch.allison2@gmail.com