Be Adventurous, a Music Recommender

By Kevin (Haowen) Yang

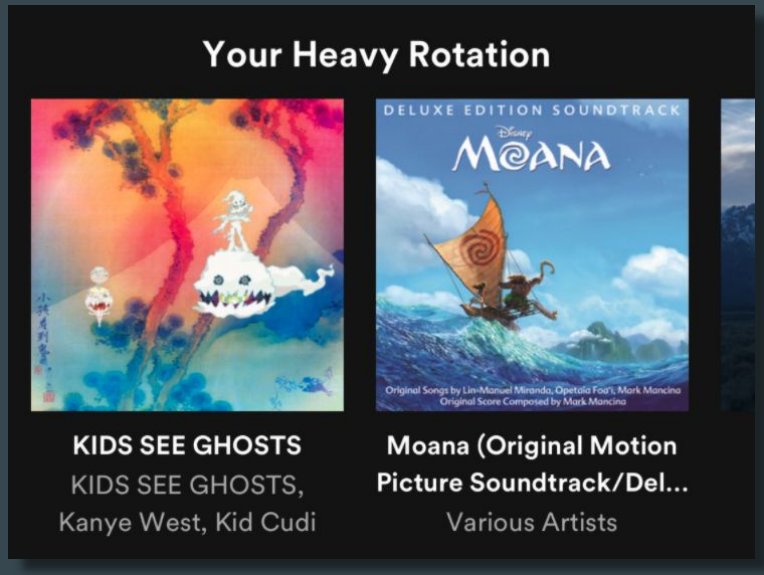
A Little About Me

- Data Scientist
- Experience in business administration and commercial bank risk management
- Graduating from Galvanize!
- Love hiking

- Love music

Background

- Americans are spending on average more than 4.5 hours a day (32.1 hours per week) listening to music, says Nielsen in releasing new data on music consumption.
- According to a survey by Deezer, 18.8 percent of respondents stated that they sought out music that was new to them on a daily basis, and 6.4 percent said that they never actively looked for new music outside their favorite genre.
- Encourage users to explore different music



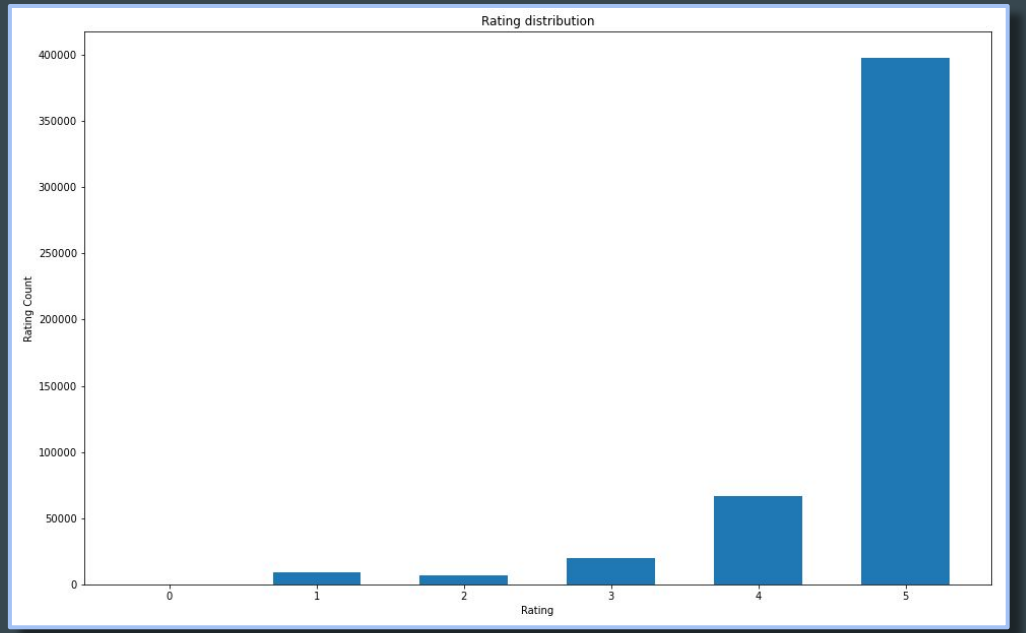
About the Dataset

- Amazon user reviews of digital music purchase provided by people from UCSD
- Original dataset contains more than 1.6 million explicit user ratings; Almost 800k users and 70k unique music products (songs and albums)
- Trimmed dataset down to 50k user ratings; Still have about 16k users and 17k music products

About the Dataset Continued

- On average, a user rates 3 songs
- 800,000 users * 70,000 songs
- That's 56 billion data points!

	Song1	Song2	Song3	Song4
User1	4	N/A	1	N/A
User2	N/A	N/A	2	N/A
User3	3	4	N/A	N/A
User4	N/A	5	N/A	N/A



Recommender Walkthrough

```
(base) Haowens-MBP:adventurous-music-recommender okugotme$ python amr_app.py
Welcome to Adventurous Music Recommender
 A music recommender that can help you explore new music
 For questions, please contact haowyang9@gmail.com. Feedback is always greatly appreciated ;-)
```

```
Please pick 5 songs you would like to rate:
-- Happy (From "Despicable Me 2")
-- Home
-- Hallelujah
-- All About That Bass
-- Uptown Funk
-- Blurred Lines [feat. T.I. & Pharrell]
-- Shake It Off
-- Roar
-- Amazing Grace
-- Radioactive
-- O Holy Night
-- Honest Face
-- Gangnam Style
```

Recommender Walkthrough Continued

```
-- Rock On
-- Chandelier
-- Hold On
-- Blown Away
-- Christmas
-- Say Something feat. Christina Aguilera
-- Gold
Enter song name here: Rolling in the Deep
Please rate (1 to 5): 3
Enter song name here:
```

```
Enter song name here: Rolling in the Deep
Please rate (1 to 5): 3
Enter song name here: All About That Bass
Please rate (1 to 5): 2
Enter song name here: When I Was Your Man
Please rate (1 to 5): 4
Enter song name here: Take Me to Church
Please rate (1 to 5): 4
Enter song name here: Gangnam Style
Please rate (1 to 5): 4

 Generating recommendations for you...
_
```

Recommender Walkthrough Continued

```
 Generating recommendations for you...
Your recommendations are:

Lost Cause
Santa Lucia
God's Grace
Glorious Day (Living He Loved Me)
Light Of The World
Enter 'r' to refresh recommendations, or 'e' to finish:
```

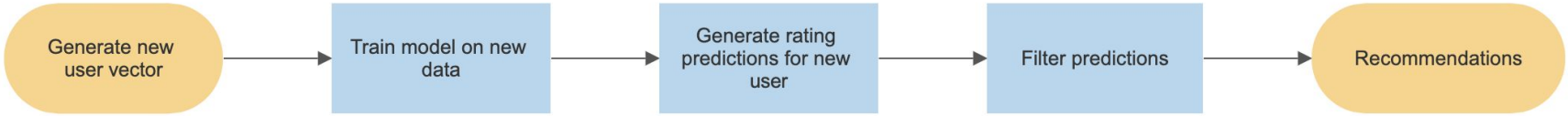
```
Afraid
I'll Always Love You
Redneck Crazy
Sirius
Something 'Bout Love
Enter 'r' to refresh recommendations, or 'e' to finish: e
Thank you for using Adventurous Music Recommender! Have a good one :-D
(base) Haowens-MBP:v1 okugotme$
```

Collaborative Filtering Model

- Surprise, a Python machine learning library
- SVDpp: a matrix factorization based algorithm
- Production model RMSE: 0.52
- Prediction vs actual ratings:

	Predicted rating	Actual rating
Song 1	4.81	5
Song 2	4.85	5
Song 3	4.48	4

Conclusion



- Generate basic recommendation with collaborative filtering model
- Filter recommendations by trimming away top 10 songs with highest predicted rating
- Refresh recommendations by sampling from recommendation list generated by previous steps

Next Steps...

- Port to Spark dataframe to scale up
- Connect to MusicBrainz database to utilize song information (i.e. performer, genre, and user-tags)

Q&A