

Million Song Recommendation system

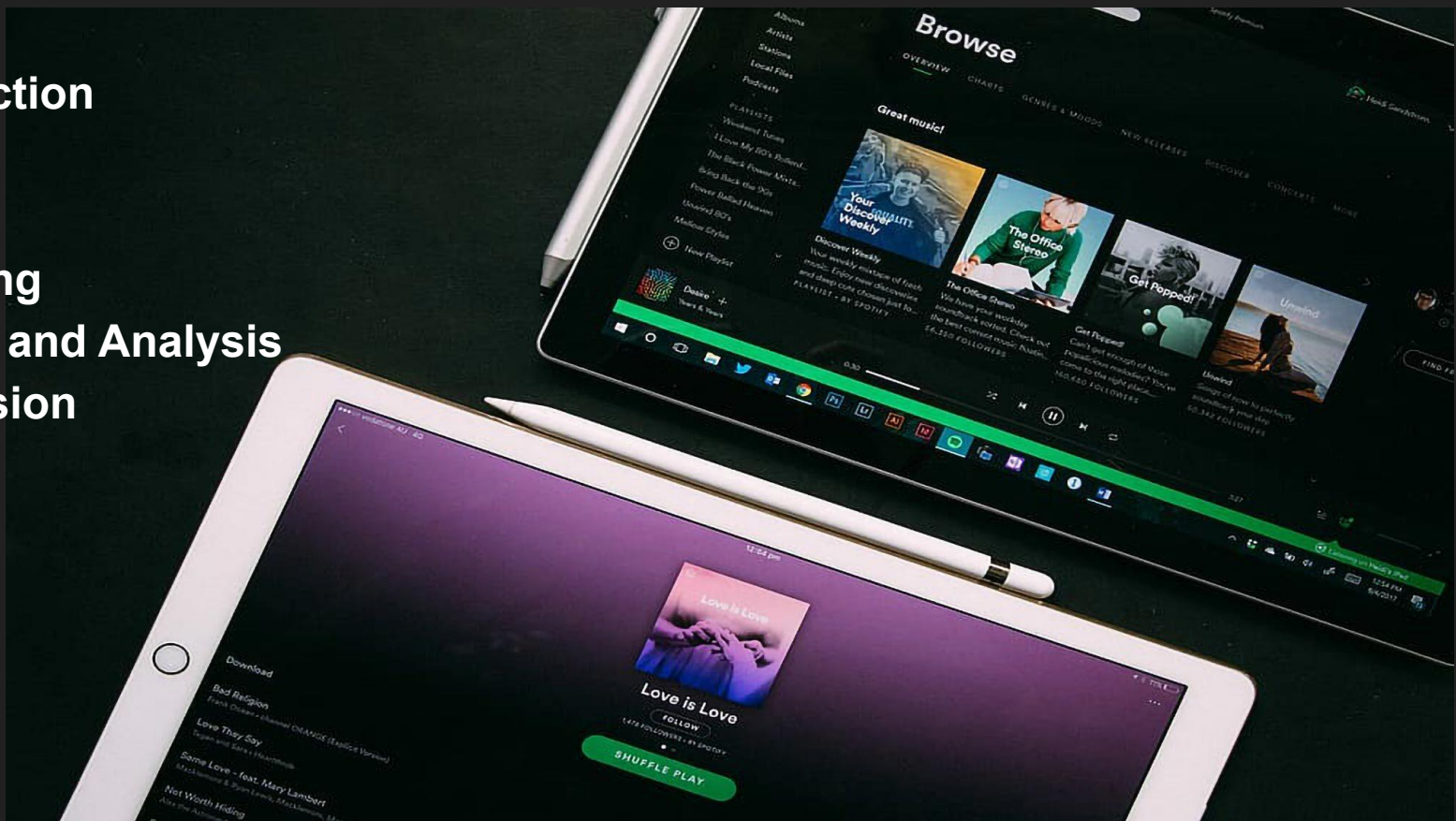
Yuen Yee Lo, PhD

Data Science Intensive Capstone Project, July 2022



Contents

- Introduction
- Data
- EDA
- Modelling
- Results and Analysis
- Conclusion



Why important?

- Recommend relevant songs a user finds on any particular platform
- Customer Satisfaction
- Personalization
- Discovery
- Provide Reports
- Increase higher users' engagement
- Increased revenue for business

Who May Care?

- Users
 - Help people to find the songs they like from million of songs
 - Explore new songs
- Business: spotify , amazon music, simux, youtube
 - Improve users engagement
 - Help prompting new songs
- Advertisement industry
- Artists, song writers

Data and Data Wrangling

Data

- Data source: <https://www.kaggle.com/competitions/msdchallenge>
- #songs: 386,213
- #User: 110,000
- Evaluation triples: 1,450,933 (song id, user id, count)

Example of data: kaggle_visible_evaluation_triplets.txt

User ID	Song ID	Count
9be82340a8b5ef32357fe5af957ccd54736ece95	SOHGGAH12A58A795BE	15
91b8fac7dc5e03f6cfaf6e2aa7171f14a8354d62	SODBXDO12A6D4FCD4F	14
91b8fac7dc5e03f6cfaf6e2aa7171f14a8354d62	SOUDGEV12A8C135FC9	10
458833ce4418010e61304b34b2c992e1cce63435	SOZVHUQ12AB0182599	6
0f40e074aab2c5f47b7ddc2277fb0295b5b3a058	SORIWEH12A8C13D4A6	54

Modelling

Cold Start Problem

- Cold start
 - New songs
 - New users
 - Low interaction (users only listen a few songs)
- Difficulty of making recommendations
- Users' feedback
 - What do they like, artists, genre, languages
 - Recommend the most popular songs of the user preference
 - lightFM model <https://github.com/lyst/lightfm>

Most popular for cold start

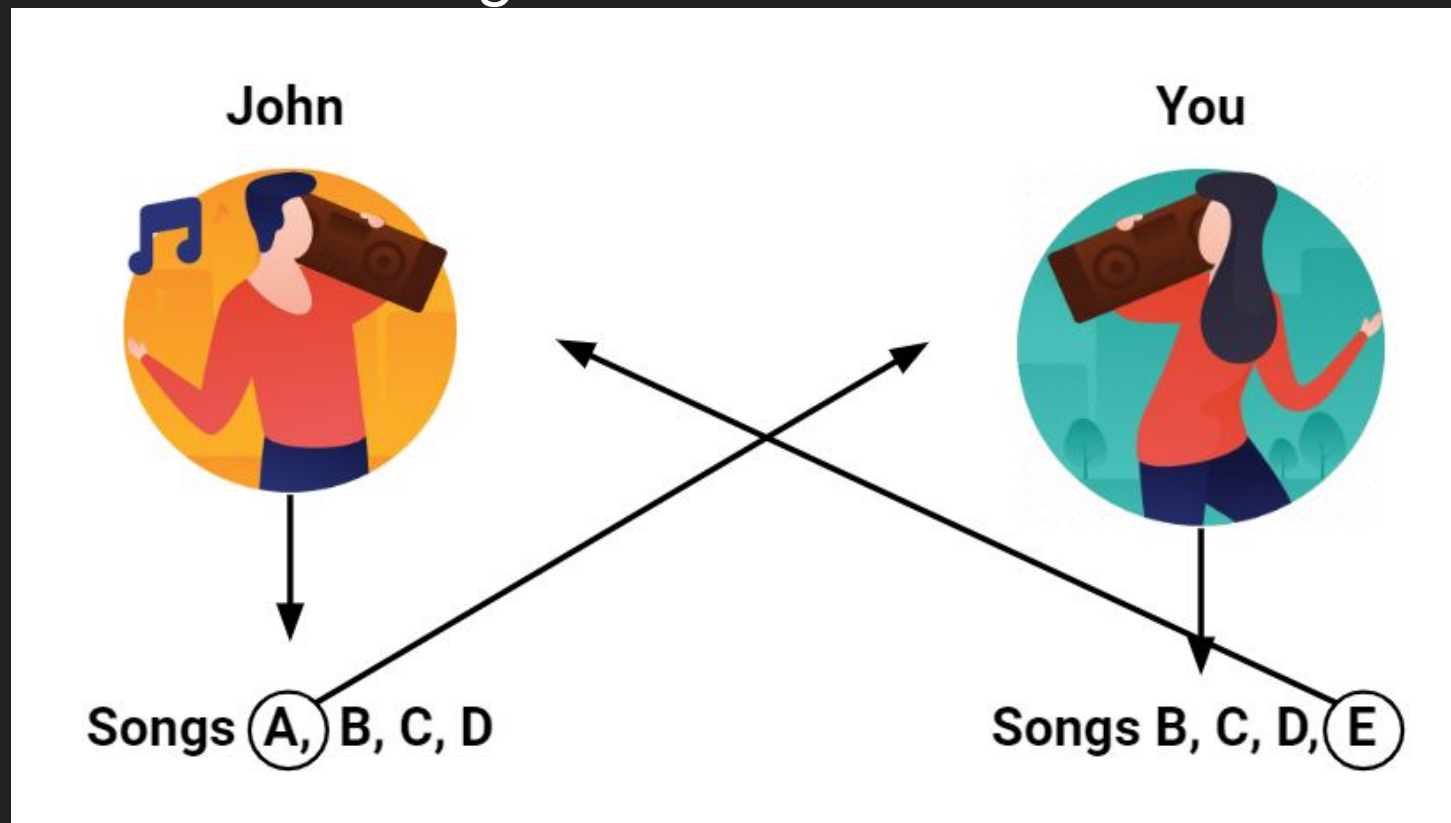
song	Total_countListened	Rank
Sehr kosmisch - Harmonia	5043	1.0
Undo - Björk	4483	2.0
You're The One - Dwight Yoakam	4136	3.0
Dog Days Are Over (Radio Edit) - Florence + Th...	3780	4.0
Revelry - Kings Of Leon	3672	5.0
Secrets - OneRepublic	3430	6.0
Horn Concerto No. 4 in E flat K495: II. Romanc...	3272	7.0
Hey_ Soul Sister - Train	2791	8.0
Fireflies - Charttraxx Karaoke	2725	9.0
Tive Sim - Cartola	2668	10.0











Similar songs: Oliver James - Fleet Foxes

	user_id	song	score	rank
0		Quiet Houses - Fleet Foxes	0.161017	1
1		Sun It Rises - Fleet Foxes	0.137725	2
2		Meadowlarks - Fleet Foxes	0.133333	3
3		He Doesn't Know Why - Fleet Foxes	0.114650	4
4		Your Protector - Fleet Foxes	0.114504	5
5		Drops In The River - Fleet Foxes	0.100000	6
6		Heard Them Stirring - Fleet Foxes	0.084906	7
7		Tiger Mountain Peasant Song - Fleet Foxes	0.073171	8
8		Ragged Wood - Fleet Foxes	0.056738	9
9		White Winter Hymnal - Fleet Foxes	0.053279	10

Collaborative Filtering

Collaborative filtering



		 Song 1	 Song 2	 Song 3	  Song 4
	User 1	✓	✗	✓	?
	User 2	✓	✓	✓	?
	User 3	✗	✗	✓	?
 	User 4	?	?	?	?

Recommendation

User Listened	Recommendation
Ballad Of Big Nothing - Elliott Smith My Paper Heart - The All-American Rejects We're All Gonna Die (featuring Iggy Pop) - Slash Saint Is A Sinner (featuring Rocco DeLuca) - Slash I Hold On (featuring Kid Rock) - Slash Alphabet Town - Elliott Smith Nothing To Say (featuring M. Shadows) - Slash Gotten (featuring Adam Levine) - Slash The Golden Rose (Album Version) - Tom Petty Genius - Kings Of Leon	Watch This (featuring Dave Grohl and Duff McKagan) - Slash Crucify The Dead (featuring Ozzy Osbourne) - Slash Starlight (featuring Myles Kennedy) - Slash Promise (featuring Chris Cornell) - Slash By The Sword (featuring Andrew Stockdale or Wolfmother) - Slash Doctor Alibi (featuring Lemmy Kilmister) - Slash The White Lady Loves You More - Elliott Smith Beautiful Dangerous (featuring Fergie) - Slash Single File - Elliott Smith Alameda - Elliott Smith

Conclusions

- Cold start:
 - Recommend most popular songs
- Item-based Collaborative Filtering
 - Recommend related songs to users

Further Improvement

- More data:
 - Song title, genre, lyrics
 - Users' playlist (see spotify Million Playlist Dataset)
- Natural language process (NLP)
 - Similarity between lyrics
 - Similarity between artists
 - Similarity between genre
- Deep learning
 - Using CNN model
 - Spectrum analysis

Thank You

Question?

Contact

Yuen Yee Lo

Email: yylo7775@gmail.com

<https://www.linkedin.com/in/yuen-yee-lo-4b74865/>

<https://github.com/yuenyeelo/>

<http://www.y2nlp.com>

Project directory: <https://github.com/yuenyeelo/springboard/tree/main/Capstone2>