


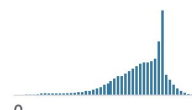
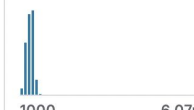
Music Recommender via Content-Based Recommendations, Matrix Factorization and Web-Scraping

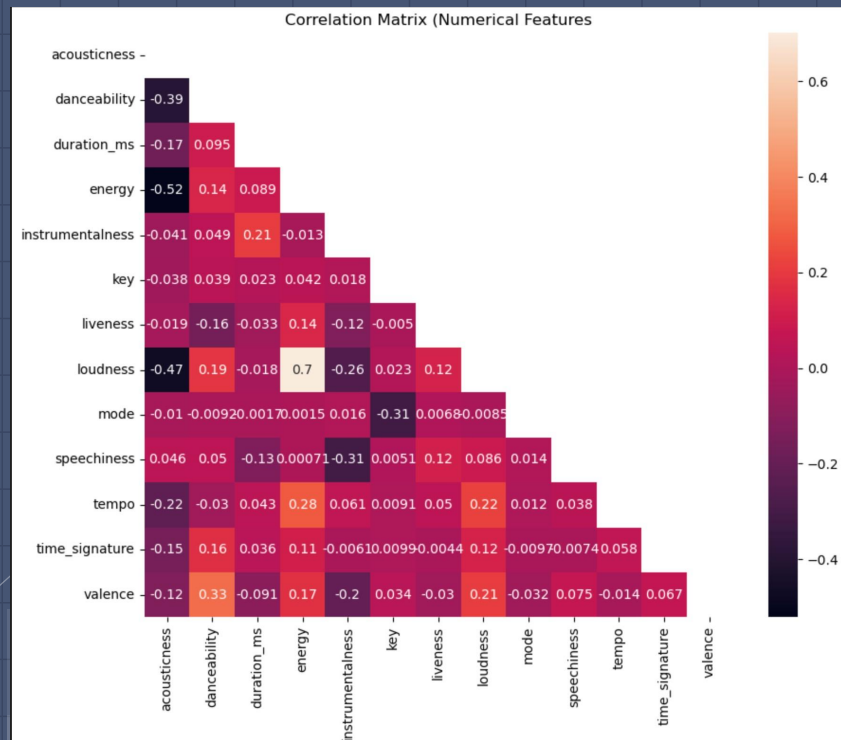
Arthur Kim, Brian Sun, Kira Jackson, Meghana Dhruv



Background & Data Source

Data Source: Audio_Features.csv file on Kaggle

isrc	# acousticness	# danceability	# duration_ms
TEXT	REAL	REAL	INTEGER
4687104 unique values			
	01	01	10006.07m
AD4X65752184	0.906	0.65	296733
AEA0D1991170	0.00095	0.621	191989
AEA0Q2004008	0.0312	0.692	282904
AEA0Q2004009	0.000558	0.516	197904
AEA0Q2004010	0.00938	0.587	199471
AEA0Q2004011	0.00854	0.656	266736



isrc - International Standard Recording Code

Methods & Analysis

Web-Scraping

- Uses Selenium chromeDriver to automate url searches
- For each ISRC code, a search is made to soundexchange.com and returns html data
- Extracts ISRC ID, artist name, song title, etc and saves into a csv file as list

ISRC
CA5KR2358074 or UPC/EAN **SEARCH**

1 recording results [CLEAR](#)

[ADD ALL TO EXPORT](#) Show releases ☐ Recordings per page 10

Artist	Title	Version	Year	Duration	ISRC
⊕ Veloz, Luis Vargas	Sexy		2023		CA5KR2358074

Export as file Refresh data Report an issue

artist title isrc

Missing: 524 (52%) Missing: 524 (52%) Missing: 0 (0%)
Distinct: 470 (47%) Distinct: 473 (47%) Distinct: 999 (100%)

470 **473** **999**
Distinct values Distinct values Distinct values

0	Missing value	Missing value	ITY701800108
1	Missing value	Missing value	US83Z1106885
2	Olivier Py / Birds of Paradise	Punk Prototype, Pt. 2	FR9W11700196
3	Newban	Find a Place to Live	GBEQ1203283
4	Father	Ghosts	BEY920901807
5	Charlotte Someone	Another Fine Day	USLZJ1956668
6	Missing value	Missing value	GBDDN1400602
7	Missing value	Missing value	NLCK41065531
8	Golden Grand	Too Club For Saks	QZPLS2147312
9	Demolish Beatz	I Use What's Left	USXQS1923227
10	Spanless	The Essence of Truth	US83Z1448772

Methods & Analysis

Content-Based Recommendation

- Uses audio features (danceability, energy, etc) to standardize and compute similarity between songs with the sigmoid kernel
- Accepts a song title as input → finds the closest matches in the dataset and ranks them by similarity scores

Recommendations based on the song: Dream of a Machine

	Title	Artist	ISRC	Danceability	Energy	Valence	Tempo	Acousticness	Liveness	Similarity Score
0	I Got It Made (Re-Recorded & Remastered)	Various Artists	USA370956824	0.666	0.751	0.85	190	0.0156	0.0999	0.907972
1	Superstars	Styles Of Beyond	US3260400034	0.505	0.865	0.14	216	0.00259	0.0563	0.906435
2	Midas	Baselinez	USA2P1680614	0.72	0.827	0.563	194	0.00468	0.248	0.905506
3	Computer Glitch	Positive Postulate Records	QZMHM2093554	0.702	0.709	0.639	190	0.0109	0.0589	0.902388
4	Dice Roll	Spiral Helix	QZNWY2352808	0.563	0.782	0.812	192	0.117	0.162	0.897396
5	BANGBAP	Bob Catt The Legend	QZ5FN2082467	0.707	0.742	0.963	180	0.116	0.0472	0.896668
6	Reggaeton Backing Track - A minor	Gene2020	QM6MZ2061016	0.686	0.735	0.735	180	0.00283	0.333	0.89225
7	Te Cogio	David El Embajador	ITJ871800178	0.724	0.893	0.94	176	0.171	0.154	0.889094
8	Super Natural	Coe	UKHY2200017	0.662	0.416	0.0397	200	0.0284	0.139	0.883567
9	Something from the Old School	DJ Tempo	GBKPL1520887	0.707	0.706	0.599	174	7.25e-05	0.0912	0.881718

Methods & Analysis

Matrix Factorization

	Title	Score
0	Stars Come Out	0.992479
1	The Void	0.992363
2	Empty Pier	0.992281
3	H.O.U.S.E	0.992152
4	Looking Back	0.992002
5	Lounge	0.991623
6	Pt. 2	0.990582
7	Don't Ever Change	0.989939
8	Distorting Substance	0.988867
9	Scream	0.987854

- Gives recommendations based upon patterns within the data
- Gives score based on implicit factors such as play count
- Scores are adjusted using BM25 weighting
- Higher Score = More Likely to be Played

Conclusion

Next Steps

Top similarity score: 0.992479

Using a combined score from matrix factorization and content based recommendation, the top 10 similar songs are recommended to the user.

Next Steps: Making a User-Friendly Interface, Incorporating Additional Data Sources (Apple Music, Pandora), Optimizing Recommendation Algorithm even further

