

Datasets:

File Name	Purpose/Content	Link
spotify_dataset.csv	Unprocessed Spotify Dataset from Kaggle	spotify_dataset.csv
tiktok.csv	Unprocessed Tiktok Dataset from Kaggle	tiktok.csv
TikTok_songs_2020.csv	Unprocessed Tiktok popular songs 2020 dataset from Kaggle	TikTok_songs_2020.csv
TikTok_songs_2021.csv	Unprocessed Tiktok popular songs 2021 dataset from Kaggle	TikTok_songs_2021.csv
TikTok_songs_2022.csv	Unprocessed Tiktok popular songs 2022 dataset from Kaggle	TikTok_songs_2022.csv
TikTok_songs_2019.csv	Unprocessed Tiktok popular songs 2019 dataset from Kaggle	TikTok_songs_2019.csv
cleaned_tt.csv	Cleaned Tiktok dataset, combination of all tiktok datasets from Kaggle	cleaned_tt.csv
cleaned_spot.csv	Cleaned Spotify dataset, used for data exploration and getting lyrics	cleaned_spot.csv
track_artists.csv	Intermediate dataset used to get spotify lyrics	track_artists.csv
spot_lyrics.json	cleaned_spot.csv combined with spotify lyrics and processed to remove non-english songs	spot_lyrics.json
spot_playlists_numeric_data.csv	Intermediate dataset containing playlist-aggregated audio features used for data exploration	spot_playlists_mean_numeric_data.csv
spotify_exploded.csv	cleaned spotify dataset exploded on playlist_uris column	spotify_explode.csv
spot_final.json	final dataset after cleaning -> removed playlists with less than 100 songs	spot_final.json
spot_final_explode.json	final dataset exploded on playlist_uri, used to test hybrid models	spot_explode_agg.json
spot_explode_agg.json	final dataset exploded with aggregated audio features of playlists, used to test hybrid models	spot_explode_agg.json
cb_initial.csv	cleaned dataset for CB models	cb_initial.csv
cf_initial.csv	userID and itemID with naive rating, used to test CF models	cf_initial.csv



cf_final.csv	userID and itemID with tiktok rating, used to test CF models	cf_final.csv
cf_final2.csv	userID and itemID with tiktok rating and timestamp, used to test CF models	cf_final2.csv

Online Notebooks:

Data preprocessing and cleaning

Hybrid DeepFM v.1

<insert folder name>

File Name	Purpose/Content	Link
BT4222 Exploratory Code.ipynb	Data processing and creation of datasets	 BT4222 Exploratory Co...
Cleaning and Preprocessing.ipynb	Data Preprocessing Cleaning Feature Engineering EDA Creation of Datasets	 Cleaning and Preproce...
Hybrid DeepFM v.1.ipynb	Running the hybrid DeepFM model	Hybrid DeepFM v.1

Offline Notebooks

Getting spotify lyrics

Folder Name	File Name	Purpose/Content
get_lyrics	get_lyrics.ipynb	Downloading lyrics of tracks
get_lyrics	lid.176.bin	
get_lyrics	.env	

Neural Collaborative Filtering

Folder Name	File Name	Purpose/Content
-------------	-----------	-----------------

collaborative_filtering	IBCF.ipynb	Item Based Collaborative Filtering Model
collaborative_filtering	LightGCN.ipynb	Light Graph Convolutional Network Model
collaborative_filtering	NCF_no_tuning.ipynb	Neural Collaborative Filtering Model without hyperparameter tuning
collaborative_filtering	NCF_w_tuning.ipynb	Neural Collaborative Filtering Model with hyperparameter tuning
collaborative_filtering	cf_initial.csv	Data for LightGCN
collaborative_filtering	cf_final2.csv	Data for IBCF and NCF

Content Based Filtering

Folder Name	File Name	Purpose/Content
content_based_filtering	CB_model.ipynb	Content Based Filtering Model with calculations for precision and recall. Contains initial evaluation and final evaluation with improvements(using combination of numerical features, tiktok score, and lyrics) within the same notebook
content_based_filtering	CB_accuracy_initial.ipynb	Content Based Filtering Model with calculation for accuracy using audio features only
content_based_filtering	CB_accuracy_final.ipynb	Content Based Filtering Model with calculation for accuracy using combination of audio features, tiktok score, and lyrics
content_based_filtering	spot_final.json	Data for Content Based Filtering Model

Hybrid Models

Folder Name	File Name	Purpose/Content
hybrid	lightFM v.1.ipynb	The lightFM model and preprocessing of data for the model (no hyperparameter tuning)
hybrid	package-list.txt	Package list for conda environment to run the code
hybrid	spot_final_explode.json	Data for the model

hybrid	spot_playlists_mean_numeric_data.csv	Data for the model
--------	--------------------------------------	--------------------