Popularity Prediction using Spotify Features

DSB 122

Dillon Diatlo

Elaine Chen

Harnish Shah

Rajashree Choudhary

March 2024

Turtle Taxi Entertainment LLC

Co-Founders:

Harnish Shah
Elaine Chen
Dillon Diatlo
Rajashree Choudhary

Chris DeepPocket

Angel Investor

At the genesis of Uber, Turtle Taxi was at its lowest point.

When money-thirsty Trillionaire CEO, Chris DeepPocket, caught word of TT's unfortunate turn of events, he performed a hostile takeover.

Now, as TT CEO and self-proclaimed Taxi King, he's forcing the crew to find which features make a song popular, so he can mass produce bad new music and become a Million - Trillionaire

Project Roadmap

- Objective
- Data Cleaning and EDA
- Feature Engineering and Selection
- Modeling and Evaluation
- Limitation
- Future Scope

Objective

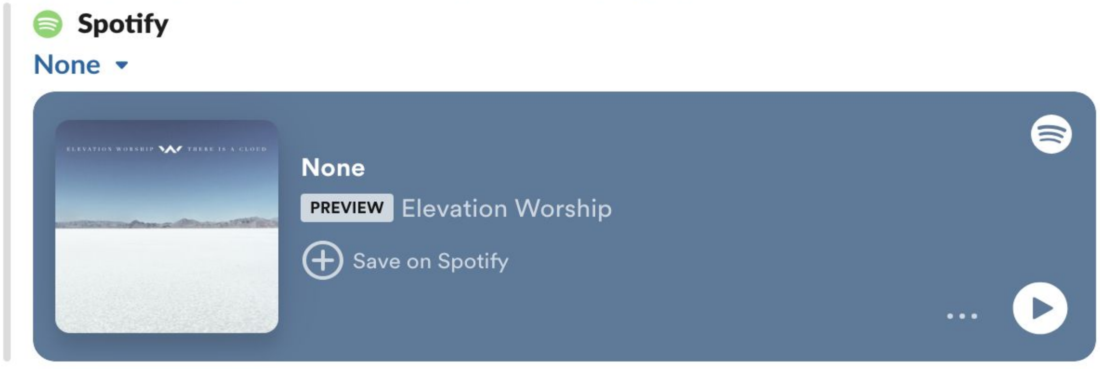
- Analyzing song features within Spotify's structure to anticipate their acclaim

- Utilizing Spotify's vast database and insights into each song's characteristics, the aim is to discern the elements contributing to a song's success.

Data Cleaning

	genre	artist_name	track_name	track_id	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	key
208285	World	Elevation Worship	NaN	7BGQCe62A58Q5ZgpQFX93t	44	0.019	0.287	350027	0.446	0.0	F

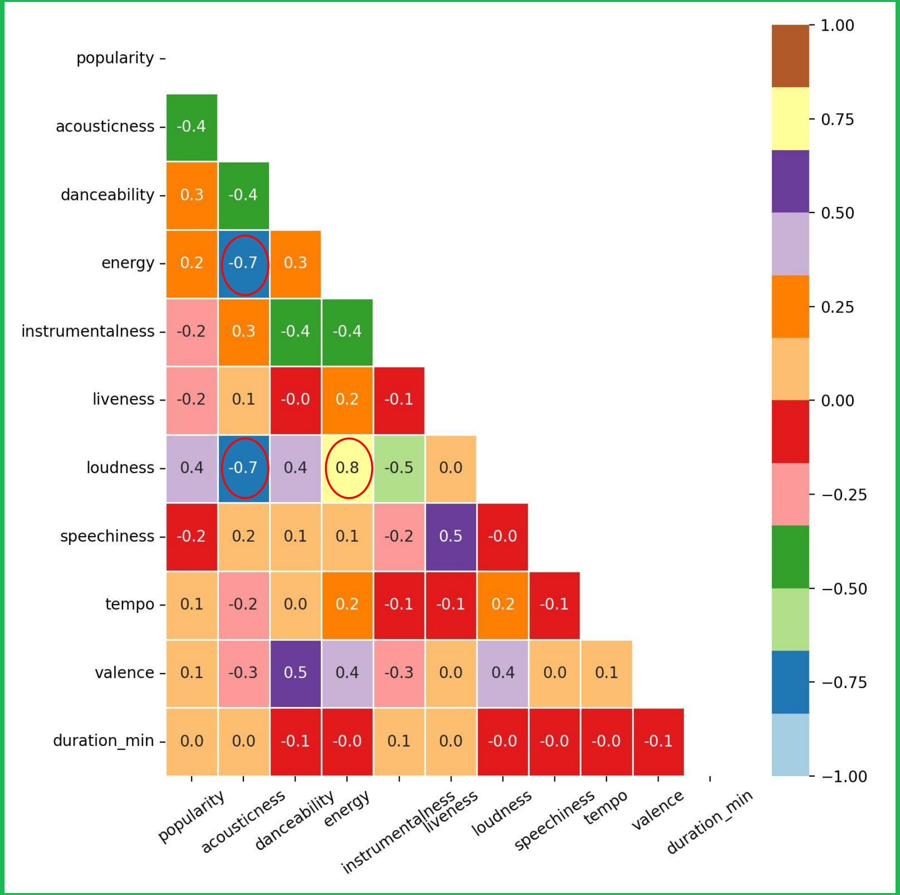
https://open.spotify.com/track/7BGQCe62A58Q5ZgpQFX93t

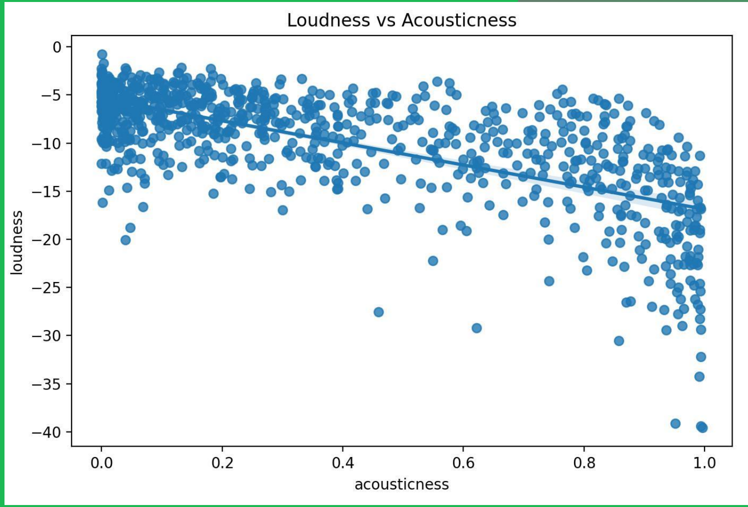


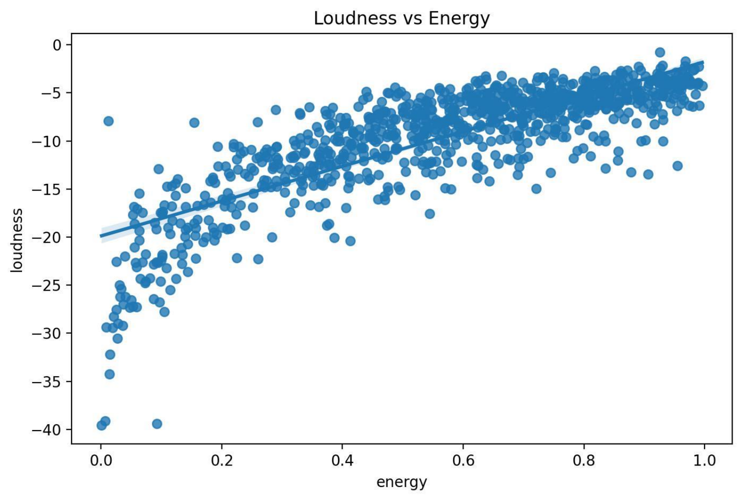
232,725
records

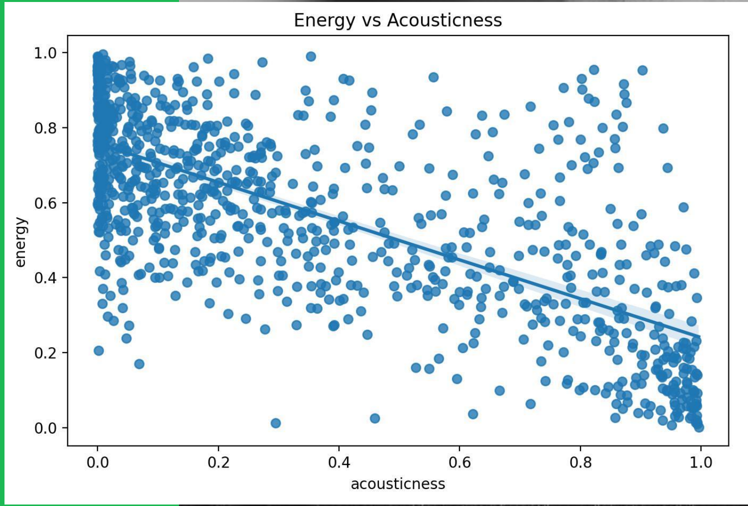
Exploratory Data Analysis

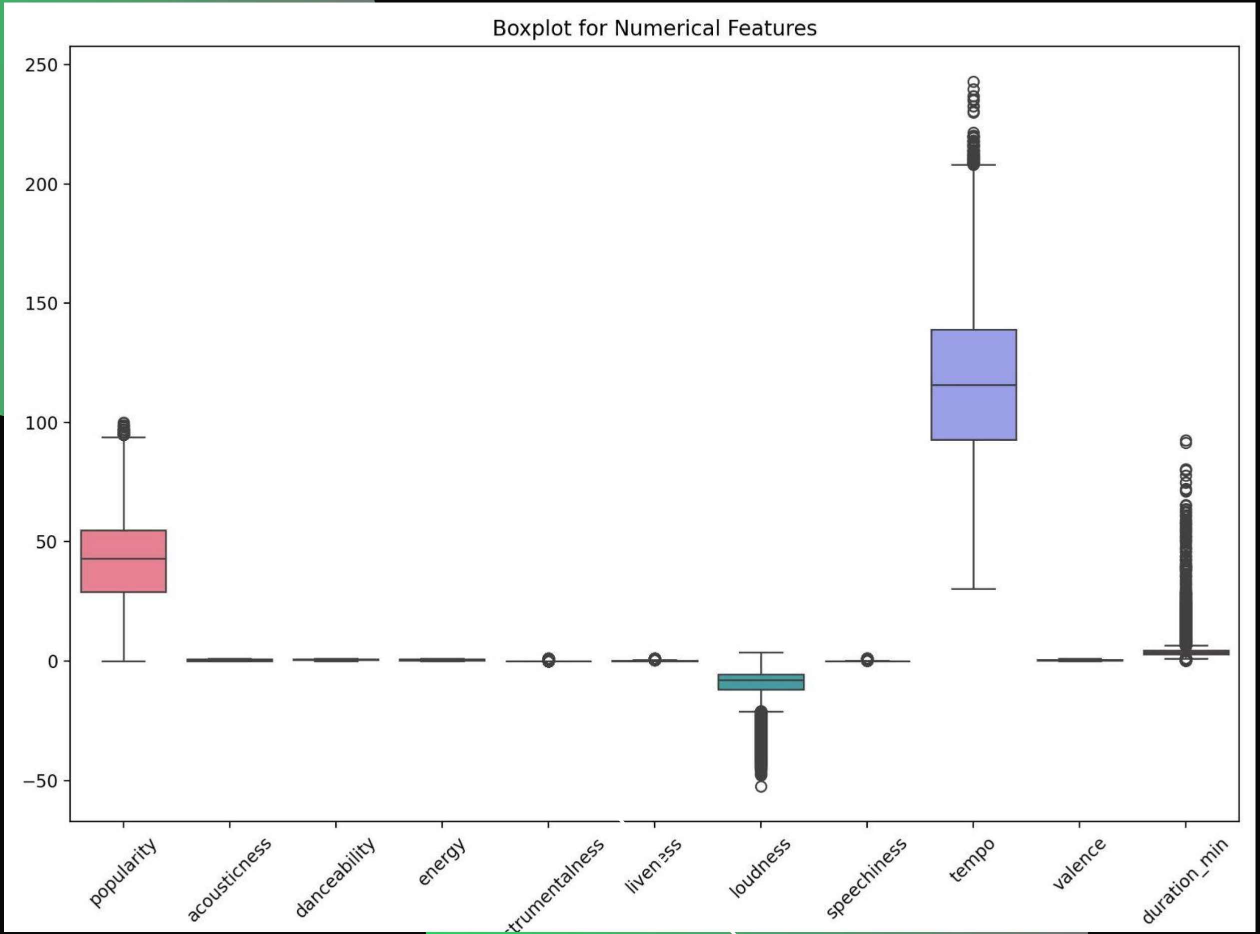
Correlation of Numeric Features




Loudness vs Acousticness
loudness
acousticness


Loudness vs Energy
loudness
energy

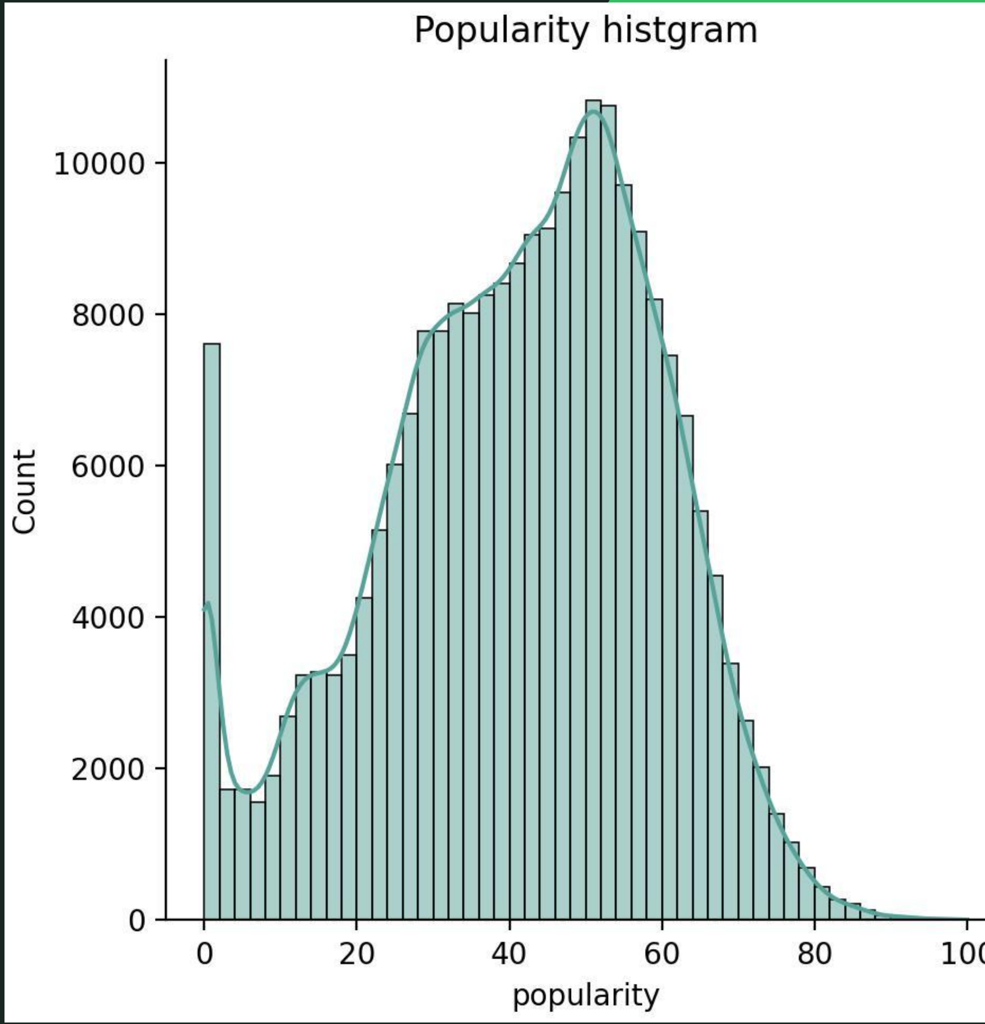

Energy vs Acousticness
energy
acousticness



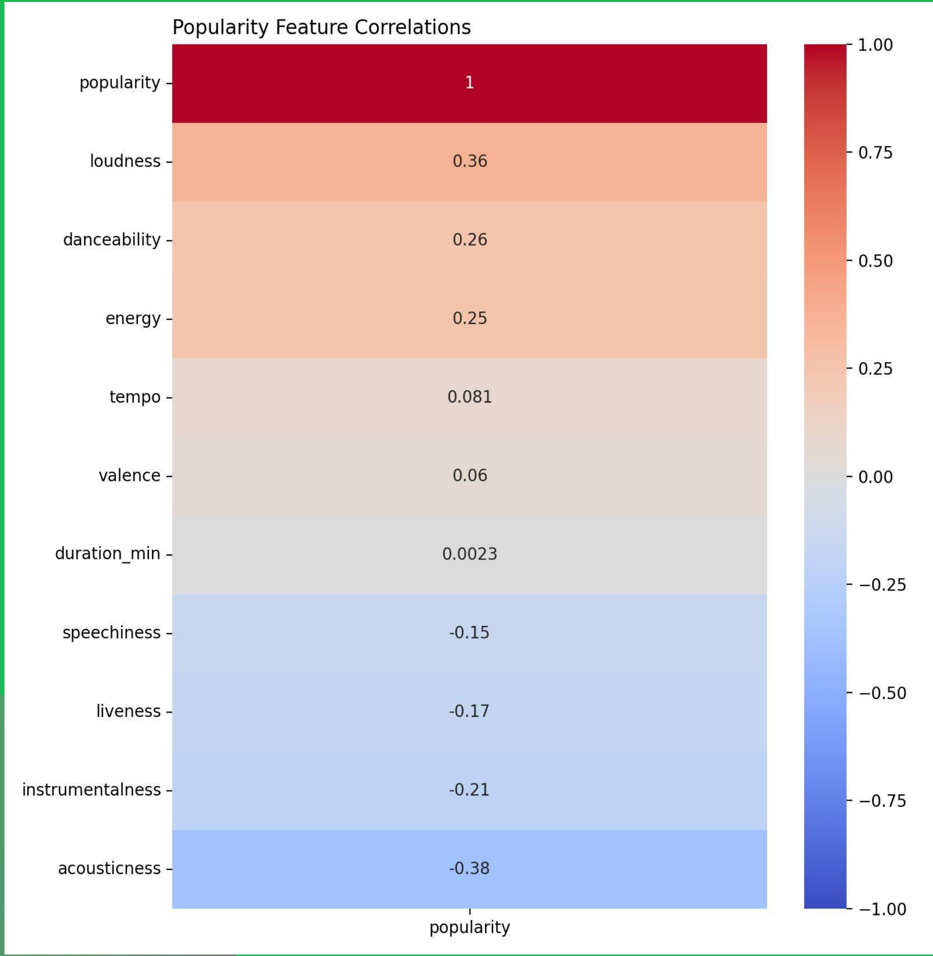
Numeric Features Statistical Summary

Target Feature:

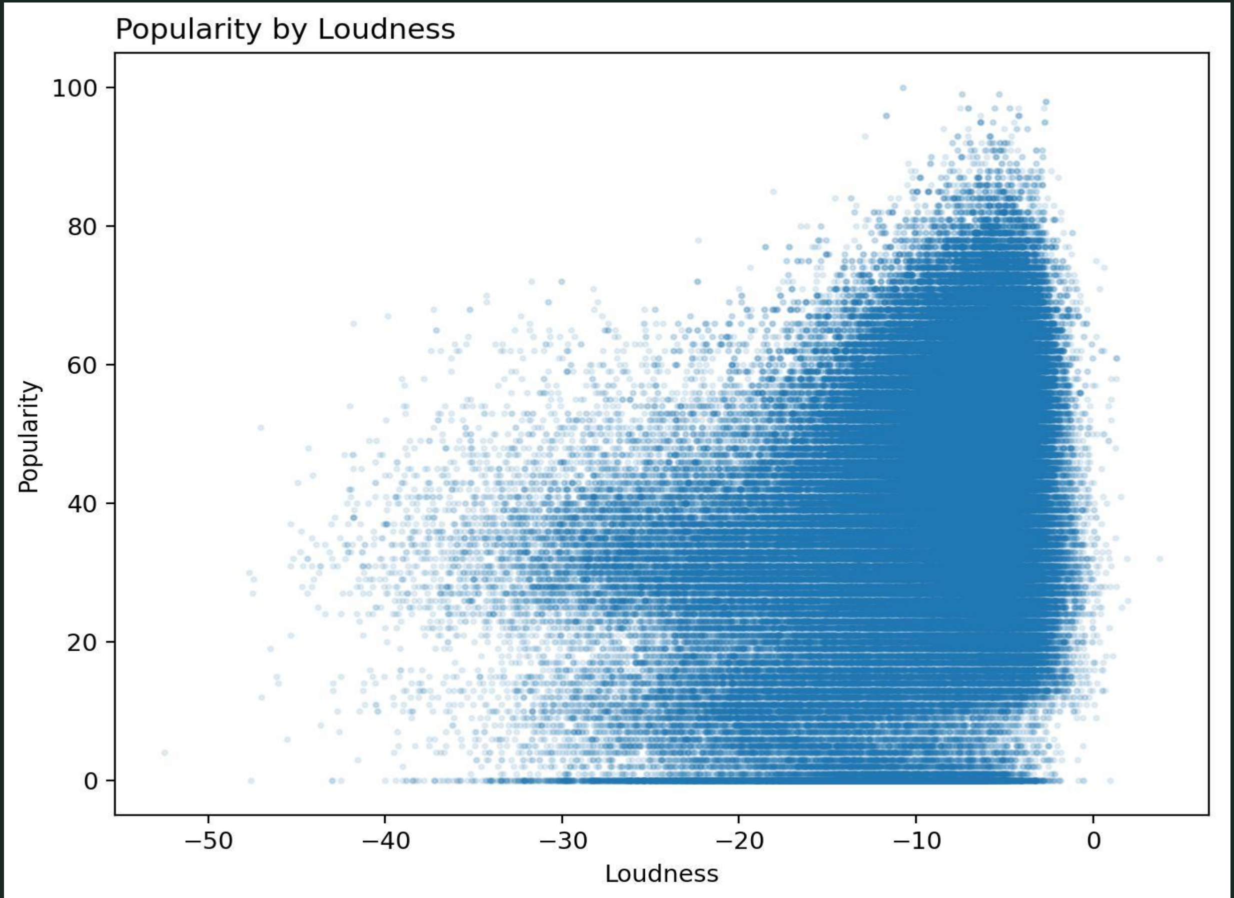
Popularity



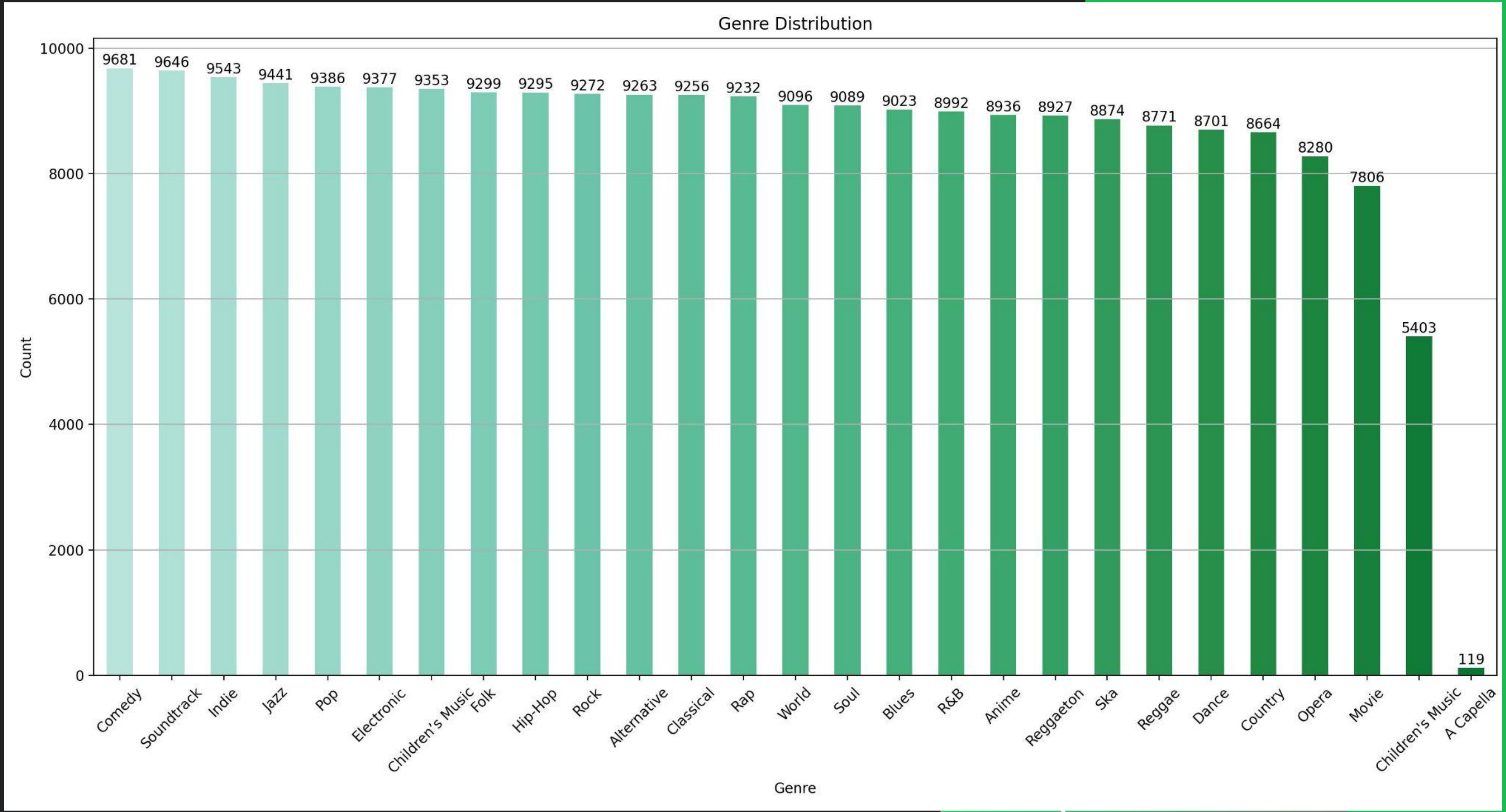
Correlation of Numeric Features with Popularity



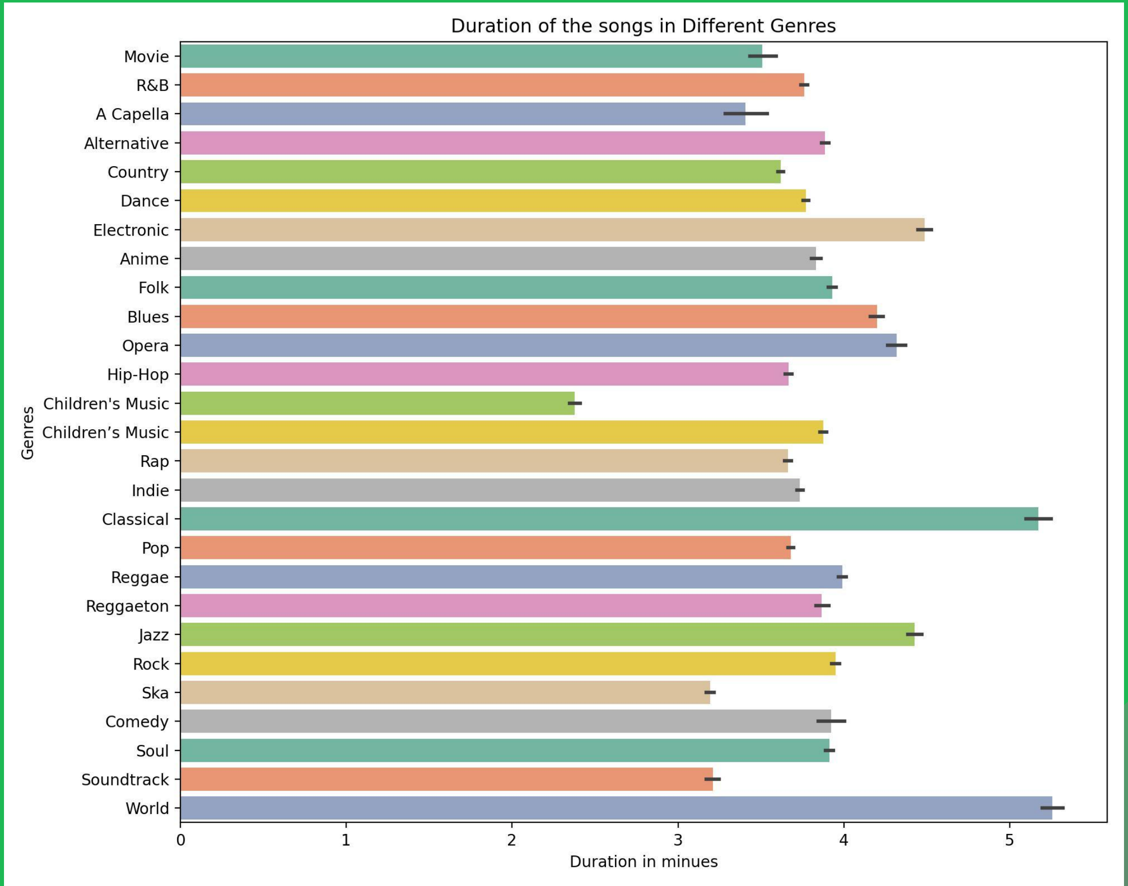
Loudness vs Popularity

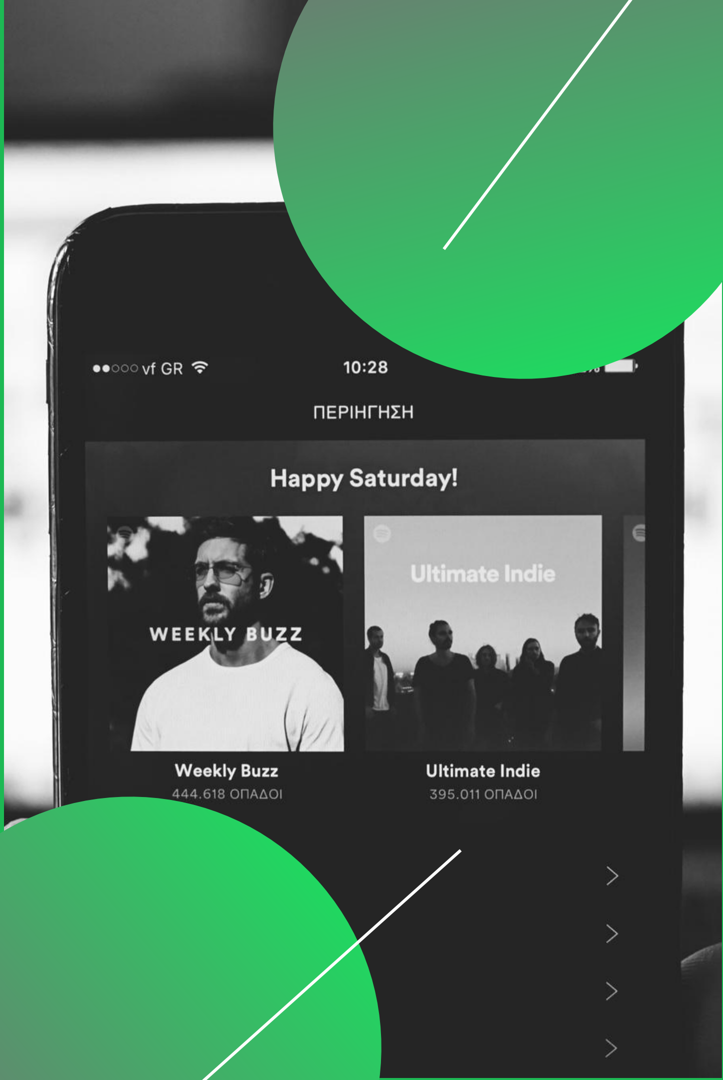


Categorical Features

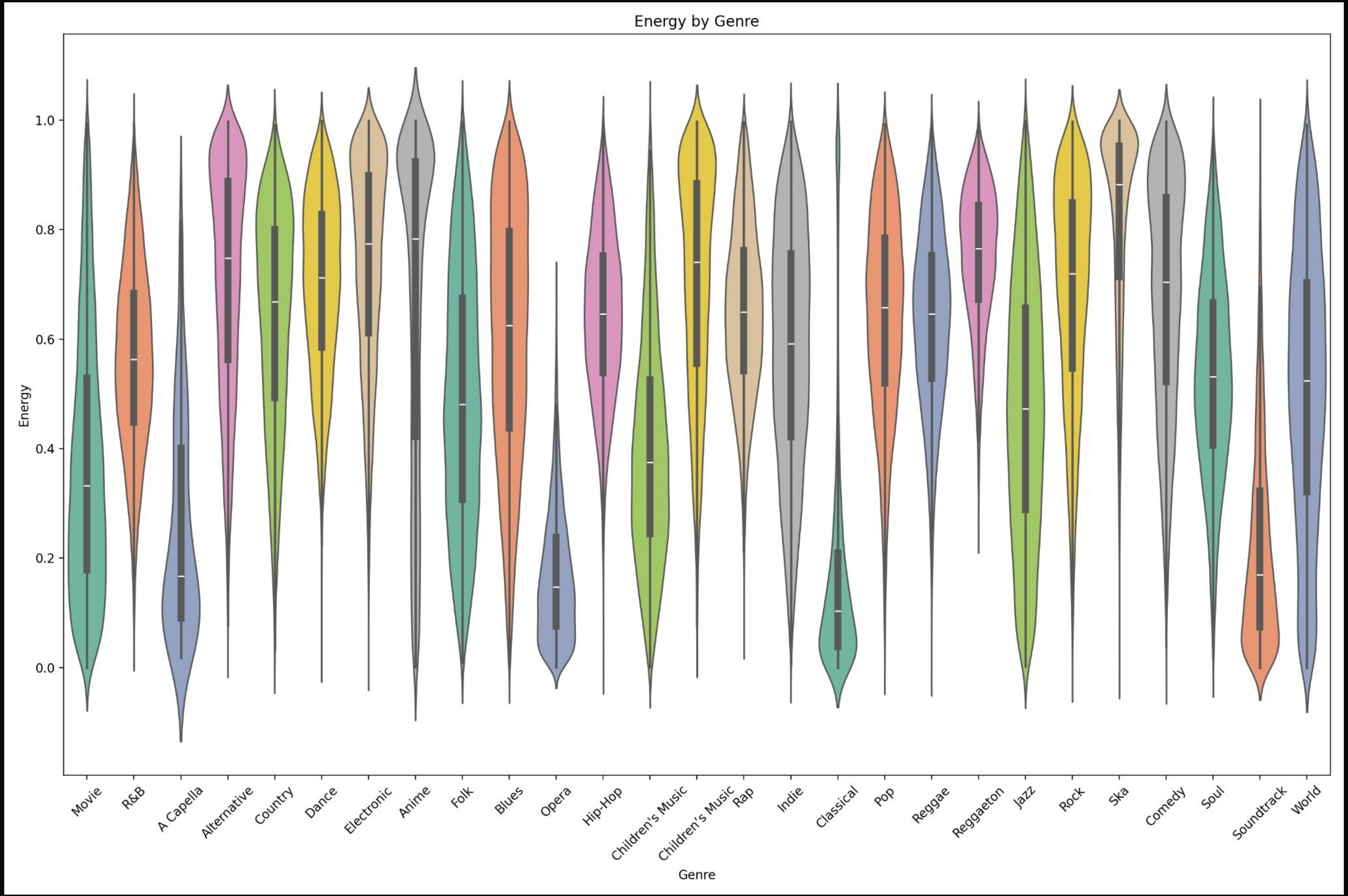


Each Genre Duration (min)

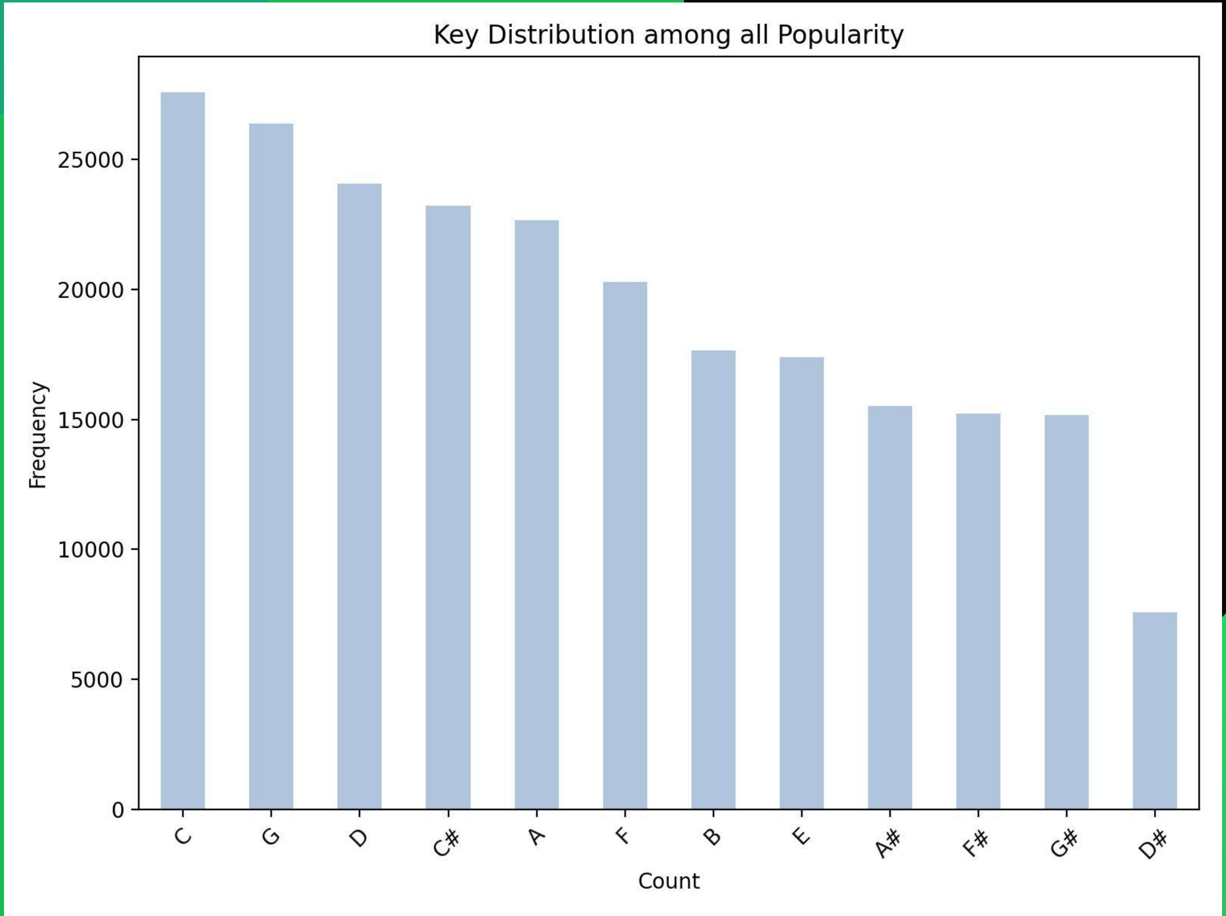




Each Genre Energy



Key Distribution



A Close look at the Data

artist_name	popularity
Drake	31703
Chris Brown	22047
Nobuo Uematsu	19710
Future	19590
Hans Zimmer	19439
Eminem	18876
Giuseppe Verdi	18580
Wolfgang Amadeus Mozart	17785
Howard Shore	17283
Johann Sebastian Bach	16508
Giacomo Puccini	16376
G-Eazy	16300
John Williams	15585
The Black Keys	15200
Frédéric Chopin	15193
Frank Ocean	14977
Mac Miller	14851
Ludwig van Beethoven	14597
Bob Marley & The Wailers	14520
J. Cole	14483

	genre	artist_name	track_name	track_id	popularity	acousticness	danceability	energy	instrumentalness	key	liveness	lo
9027	Dance	Ariana Grande	7 rings	14msK75pk3pA33pzPVNtBF	100	0.5780	0.725	0.321	0.000000	C#	0.0884	
107804	Pop	Ariana Grande	7 rings	14msK75pk3pA33pzPVNtBF	100	0.5780	0.725	0.321	0.000000	C#	0.0884	
86951	Rap	Post Malone	Wow.	6MWtB6iiXyIwun0YzU6DFP	99	0.1630	0.833	0.539	0.000002	B	0.1010	
107803	Pop	Post Malone	Wow.	6MWtB6iiXyIwun0YzU6DFP	99	0.1630	0.833	0.539	0.000002	B	0.1010	
107802	Pop	Ariana Grande	break up with your girlfriend, i'm bored	4kV4N9D1iKVxx1KLvtTpjS	99	0.0421	0.726	0.554	0.000000	F	0.1060	
9026	Dance	Ariana Grande	break up with your girlfriend, i'm bored	4kV4N9D1iKVxx1KLvtTpjS	99	0.0421	0.726	0.554	0.000000	F	0.1060	
66643	Hip-Hop	Daddy Yankee	Con Calma	5w9c2J52mkdntKOmRLeM2m	98	0.1100	0.737	0.860	0.000002	G#	0.0574	
107909	Pop	Daddy Yankee	Con Calma	5w9c2J52mkdntKOmRLeM2m	98	0.1100	0.737	0.860	0.000002	G#	0.0574	
138918	Reggaeton	Daddy Yankee	Con Calma	5w9c2J52mkdntKOmRLeM2m	98	0.1100	0.737	0.860	0.000002	G#	0.0574	
107829	Pop	Ava Max	Sweet but Psycho	25sgk305KZfyuqVBQlahim	97	0.0691	0.719	0.704	0.000000	C#	0.1660	

Feature Engineering and Selection

FEATURE ENGINEERING

Column Genre:
Each genre in each column as binary (0,1)

Map column Mode

Column Key:
turn to a float feature

Column Duration:
turn unit into minute

FEATURE SELECTION

Drop track ID

Track name

Artists name

Duration

Modeling and Evaluation

MODEL	TYPE	PARAMS	SCORES	
K-Means Cluster	Unsupervised	rang(2, 11) n_clusters=k	Best k = 3 Sil = 0.30	
3 CLASSES				
DNN1	Supervised	Dense(100, relu, …) Dense(50, relu) Dense(3, act='softmax')	79% Val Accuracy	42.9% Val Loss
DNN 2	Supervised	Dense(100, relu) Dropout(0.5) Dense(50, relu) Dropout(0.5) Dense(65, relu) Dropout(0.5) Dense(3, act='softmax')) } x2	57% Val Accuracy	91% Val Loss
RandomForest	Supervised	Default	99% Train	70% Test

MODEL	TYPE	PARAMS	SCORES	
5 CLASSES				
DNN1	Supervised	Dense(100, relu, …) Dense(50, relu) Dense(5, act='softmax')	73%	73%
DNN2	Supervised	Dense(100, relu) Dropout(0.5) Dense(50, relu) Dropout(0.5) Dense(65, relu) Dropout(0.5) Dense(5, act='softmax')) } x2	57%	58%
RandomForest	Supervised	Default	77% Train	55% Test

79% acc

DENSE NEURAL NETWORK

Baseline:

0	0.393900
1	0.567331
2	0.038769

LIMITATIONS

1. Out-of-date Data

2. Class Imbalance

Future Scope

UPDATE

Prioritize finding more up-to-date data that includes more features, such as datetime and streams.

BALANCE

Implement measures to more evenly balance classes to ensure robust analysis and accurate predictions.

COLLABORATE

Connect with someone from the Spotify team.

TUNE

Continue to tune and tweak parameters to achieve more accurate predictions.

Thank You

Do you have any question?

Elaine Chen

Harnish Shah

Dillon Diatlo

Rajashree Choudhary