MUSIC Recommendation using Spotify Dataset

ZIKI

HOME

CONTENTS

PLAYLIST

01 *Intro*

02 *Data*

03 *Model*

04 *Result*

THANKS!

Contents

Mars Is a Cold Place
The 15th Planet

2:54 3:49

01 Introduction

Mars Is a Cold Place
The 15th Planet

2:54 3:49

Introduction

The main goal is to recommend songs based on their unique features and share insights about the most popular tracks with a wider audience.

Using machine learning, we want to present users with song suggestions that align with specific musical characteristics.

02

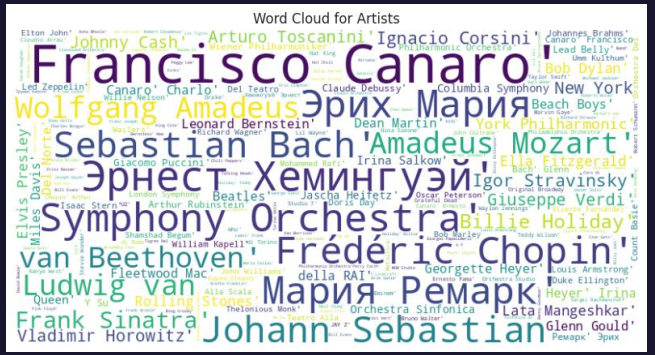
Data Engineering

```
Data columns (total 19 columns):
 #   Column            Non-Null Count   Dtype
---  ------            --------------   -----
 0   valence           170653 non-null  float64
 1   year              170653 non-null  int64
 2   acousticness      170653 non-null  float64
 3   artists           170653 non-null  object
 4   danceability      170653 non-null  float64
 5   duration_ms       170653 non-null  int64
 6   energy            170653 non-null  float64
 7   explicit          170653 non-null  int64
 8   id                170653 non-null  object
 9   instrumentalness  170653 non-null  float64
 10  key               170653 non-null  int64
 11  liveness          170653 non-null  float64
 12  loudness          170653 non-null  float64
 13  mode              170653 non-null  int64
 14  name              170653 non-null  object
 15  popularity        170653 non-null  int64
 16  release_date      170653 non-null  object
 17  speechiness       170653 non-null  float64
 18  tempo             170653 non-null  float64
dtypes: float64(9), int64(6), object(4)
```

Data Collection

A Spotify dataset has been collected from **Kaggle**, which includes parameters like acousticness, artists, danceability, loudness, popularity,etc.

The dataset consists of **170653 rows and 19 columns**



Mars Is a Cold Place
The 15th Planet
2:54
3:49

	valence	year	acousticness	artists	danceability	duration_ms	energy	explicit	id	instrumentalness	key	liveness	loudness	mode	name	popularity	release_date	speechiness	tempo
0	0.0594	1921	0.98200	['Sergei Rachmaninoff', 'James Levine', 'Berli...	0.279	831667	0.211	0	4BJqT0PrAfrxzMOxytFOIz	0.878000	10	0.6650	-20.096	1	Piano Concerto No. 3 in D Minor, Op. 30: III. ...	4	1921	0.0366	80.954
1	0.9630	1921	0.73200	['Dennis Day']	0.819	180533	0.341	0	7xPhfUan2yNtyFG0cUWkt8	0.000000	7	0.1600	-12.441	1	Clancy Lowered the Boom	5	1921	0.4150	60.936
2	0.0394	1921	0.96100	['KHP Kridhamardawa Karaton Ngayogyakarta Hadi...	0.328	500062	0.166	0	1o6I8BglA6ylDMrIELygv1	0.913000	3	0.1010	-14.850	1	Gati Bali	5	1921	0.0339	110.339
3	0.1650	1921	0.96700	['Frank Parker']	0.275	210000	0.309	0	3ftBPsC5vPBKxYSee08FDH	0.000028	5	0.3810	-9.316	1	Danny Boy	3	1921	0.0354	100.109
4	0.2530	1921	0.95700	['Phil Regan']	0.418	166693	0.193	0	4d6HGyGT8e121BsdKmw9v6	0.000002	3	0.2290	-10.096	1	When Irish Eyes Are Smiling	2	1921	0.0380	101.665
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
170648	0.6080	2020	0.08460	['Anuel AA', 'Daddy Yankee', 'KAROL G', 'Ozuna...	0.786	301714	0.808	0	0KkIkfsLEJbrcIhYsCL7L5	0.000289	7	0.0822	-3.702	1	China	72	2020	0.0881	105.029
170649	0.7340	2020	0.20600	['Ashnikko']	0.717	150654	0.753	0	0OStKKAuXlxA0fMH54Qs6E	0.000000	7	0.1010	-6.020	1	Halloweenie III: Seven Days	68	2020	0.0605	137.936
170650	0.6370	2020	0.10100	['MAMAMOO']	0.634	211280	0.858	0	4BZXVFYCb76Q0Klojq4piV	0.000009	4	0.2580	-2.226	0	AYA	76	2020	0.0809	91.688
170651	0.1950	2020	0.00998	['Eminem']	0.671	337147	0.623	1	5SiZJoLXp3WOl3J4C8IK0d	0.000008	2	0.6430	-7.161	1	Darkness	70	2020	0.3080	75.055
170652	0.6420	2020	0.13200	['KEVVO', 'J Balvin']	0.856	189507	0.721	1	7HmnJHfs0BkFzX4x8j0hkI	0.004710	7	0.1820	-4.928	1	Billetes Azules (with J Balvin)	74	2020	0.1080	94.991

170653 rows × 19 columns

Exploratory Data Analysis (EDA)

- Histagram
- Correlation Matrix
- Distribution of popularity

HOME

CONTENTS

PLAYLIST

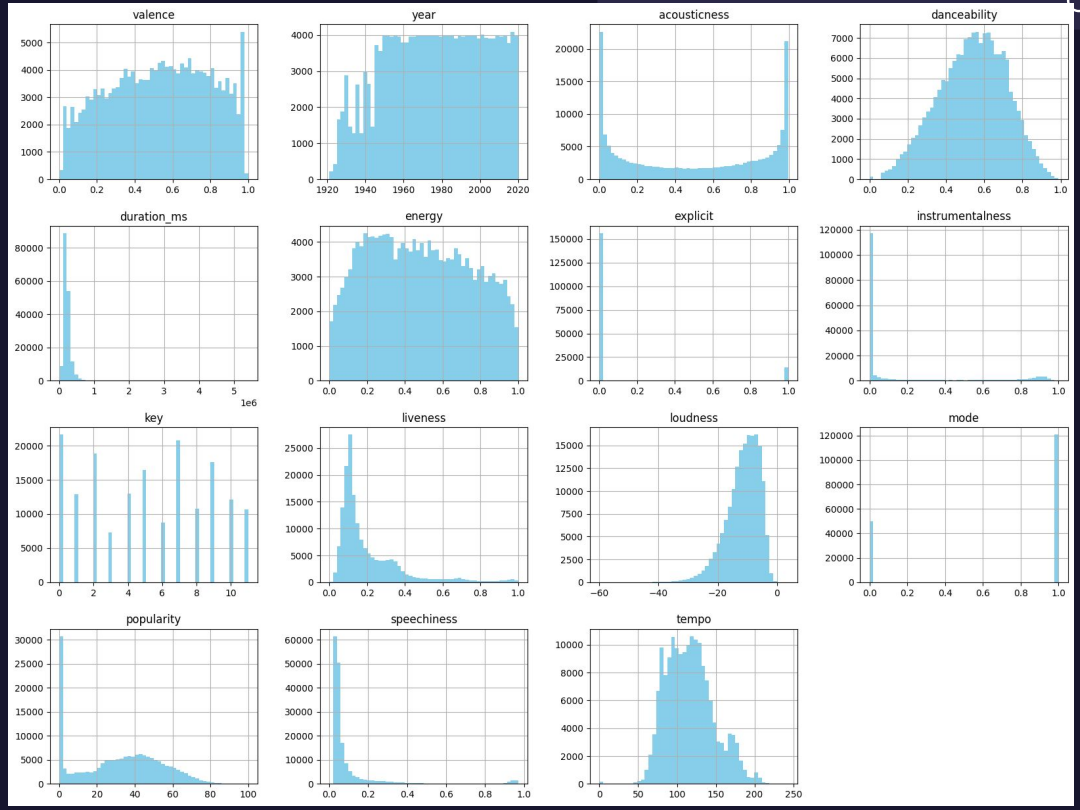
01 *Intro*

02 *Data*

03 *Model*

04 *Result*

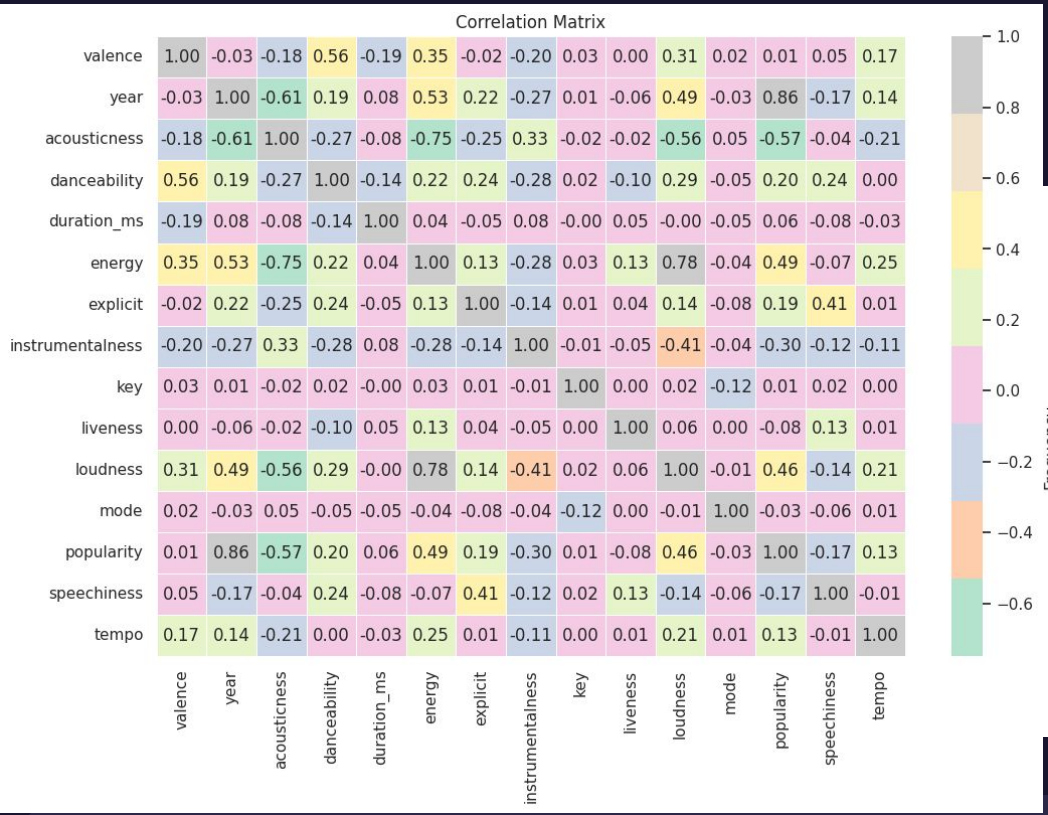
THANKS!

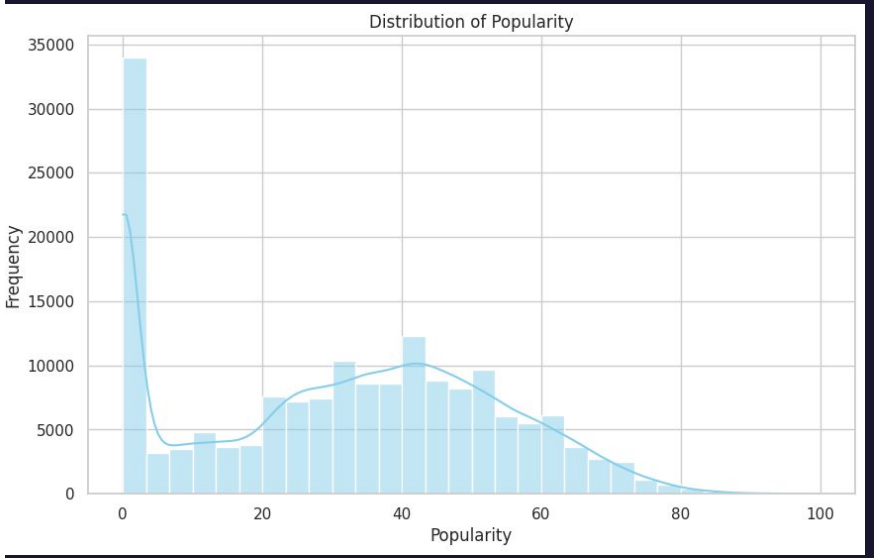


Mars Is a Cold Place
The 15th Planet

2:54

3:49


Correlation Matrix
valence 1.00 -0.03 -0.18 0.56 -0.19 0.35 -0.02 -0.20 0.03 0.00 0.31 0.02 0.01 0.05 0.17
year -0.03 1.00 -0.61 0.19 0.08 0.53 0.22 -0.27 0.01 -0.06 0.49 -0.03 0.86 -0.17 0.14
acousticness -0.18 -0.61 1.00 -0.27 -0.08 -0.75 -0.25 0.33 -0.02 -0.02 -0.56 0.05 -0.57 -0.04 -0.21
danceability 0.56 0.19 -0.27 1.00 -0.14 0.22 0.24 -0.28 0.02 -0.10 0.29 -0.05 0.20 0.24 0.00
duration_ms -0.19 0.08 -0.08 -0.14 1.00 0.04 -0.05 0.08 -0.00 0.05 -0.00 -0.05 0.06 -0.08 -0.03
energy 0.35 0.53 -0.75 0.22 0.04 1.00 0.13 -0.28 0.03 0.13 0.78 -0.04 0.49 -0.07 0.25
explicit -0.02 0.22 -0.25 0.24 -0.05 0.13 1.00 -0.14 0.01 0.04 0.14 -0.08 0.19 0.41 0.01
instrumentalness -0.20 -0.27 0.33 -0.28 0.08 -0.28 -0.14 1.00 -0.01 -0.05 -0.41 -0.04 -0.30 -0.12 -0.11
key 0.03 0.01 -0.02 0.02 -0.00 0.03 0.01 -0.01 1.00 0.00 0.02 -0.12 0.01 0.02 0.00
liveness 0.00 -0.06 -0.02 -0.10 0.05 0.13 0.04 -0.05 0.00 1.00 0.06 0.00 -0.08 0.13 0.01
loudness 0.31 0.49 -0.56 0.29 -0.00 0.78 0.14 -0.41 0.02 0.06 1.00 -0.01 0.46 -0.14 0.21
mode 0.02 -0.03 0.05 -0.05 -0.05 -0.04 -0.08 -0.04 -0.12 0.00 -0.01 1.00 -0.03 -0.06 0.01
popularity 0.01 0.86 -0.57 0.20 0.06 0.49 0.19 -0.30 0.01 -0.08 0.46 -0.03 1.00 -0.17 0.13
speechiness 0.05 -0.17 -0.04 0.24 -0.08 -0.07 0.41 -0.12 0.02 0.13 -0.14 -0.06 -0.17 1.00 -0.01
tempo 0.17 0.14 -0.21 0.00 -0.03 0.25 0.01 -0.11 0.00 0.01 0.21 0.01 0.13 -0.01 1.00
valence year acousticness danceability duration_ms energy explicit instrumentalness key liveness loudness mode popularity speechiness tempo
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4 −0.6

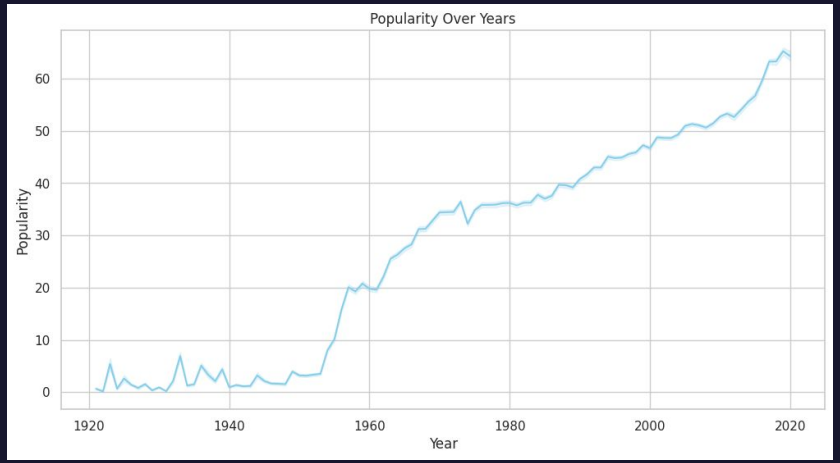

Distribution of Popularity
Frequency
35000
30000
25000
20000
15000
10000
5000
0
0
20
40
60
80
100
Popularity

Data Preprocessing

Dropping unnecessary columns:
'artists', 'id', 'name', 'release_date','Year'

Data Splitting: 80% Training & 20% Testing

Creating preprocessing pipelines for 12 numerical attributes :
"acousticness", "danceability", "duration_ms", "energy", "explicit", "instrumentalness", "key", "liveness", "loudness", "mode", "speechiness", "tempo"



03 Model Development

Mars Is a Cold Place
The 15th Planet

2:54 3:49

Models Explored

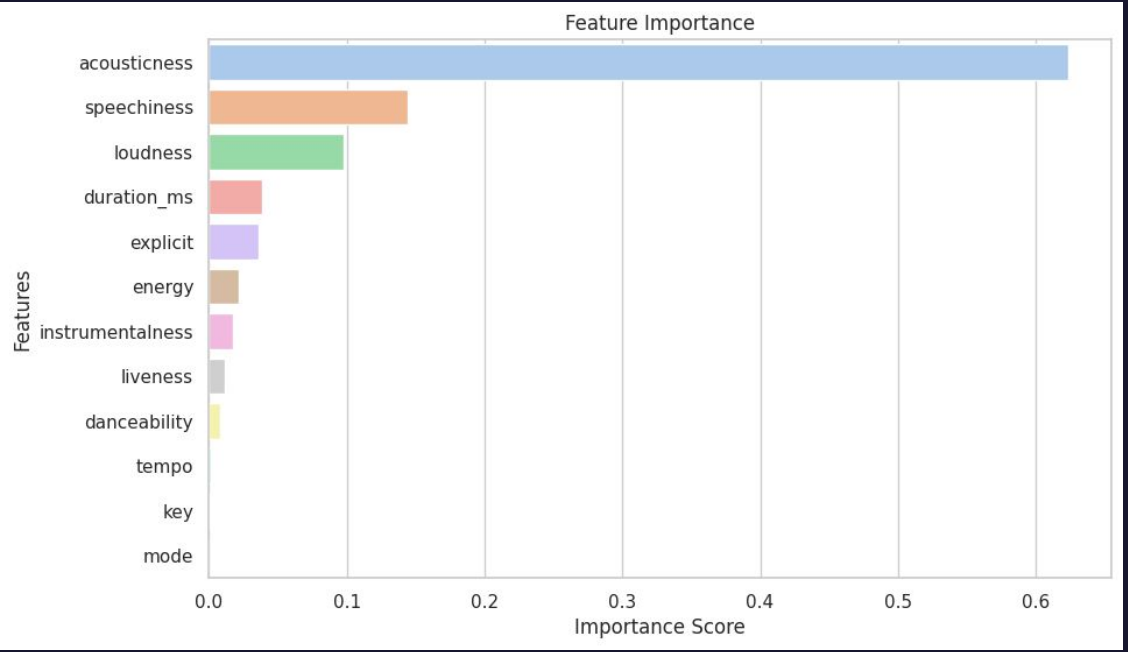
- Linear Regression
- Polynomial Regression (2nd and 3rd degree)
- Support Vector Machines (Linear and 2nd-degree Polynomial)
- Decision Tree
- Random Forest

04 Results & Conclusion

Results

	Train RMSE	Test RMSE
Linear Regression	16.619	16.617
Polynomial Regression(2nd)	15.231	15.185
Polynomial Regression(3rd)	14.706	14.761
SVM	16.965	16.970
Polynomial SVM(2nd)	18.550	18.549
Decision Tree	14.679	14.874
Random Forest	14.421	14.582

Feature Importance



Mars Is a Cold Place
The 15th Planet

2:54

3:49

Conclusion

In summary, this project assessed the predictive performance of seven models using RMSE on training and testing datasets.

In conclusion, all models performed reasonably well, with RMSE ranging from 14 to 18.55. The **Random Forest** model stood out with the lowest RMSE,(Train RMSE: 14.42, Test RMSE: 14.58) while the **Second-degree SVM** exhibited the highest(Train and Test RMSE: 18.55).

Future Innovation

- Suggest the most popular songs in real-time
- Add personalized recommendations based on users‘ listening history

This combines **personalized song suggestions with popular tracks** that a lot of people like. It will give music lovers a complete experience.

Thanks!