Music Recommendation

Enoch Kwaning

Introduction

With the rise of digital content distribution, we now have access to huge music collections. With millions of songs to choose from, we sometimes tend to get overwhelmed with finding what to listen to. Thus, an efficient music recommender system is necessary.

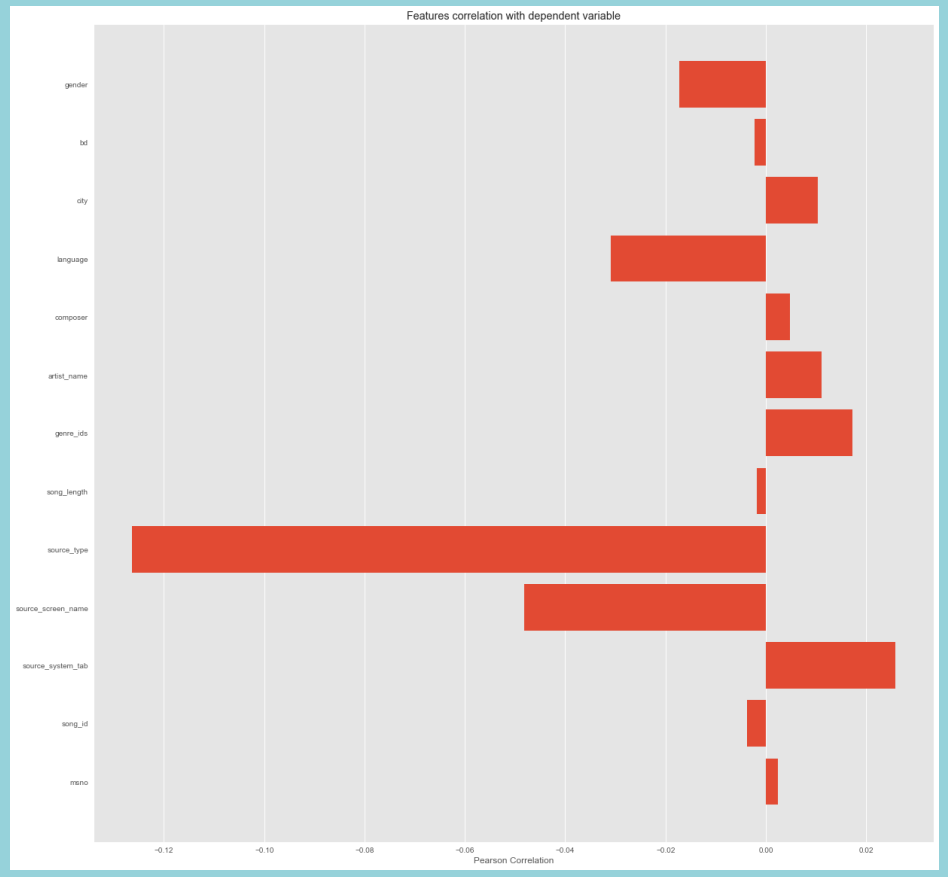
For this project we will be comparing two different approaches to music recommendation and finding the most efficient.

The KKbox dataset with the target variable being whether the user has had **recurring listening events within a month**

The Spotify Dataset with the target variable being **popularity.**

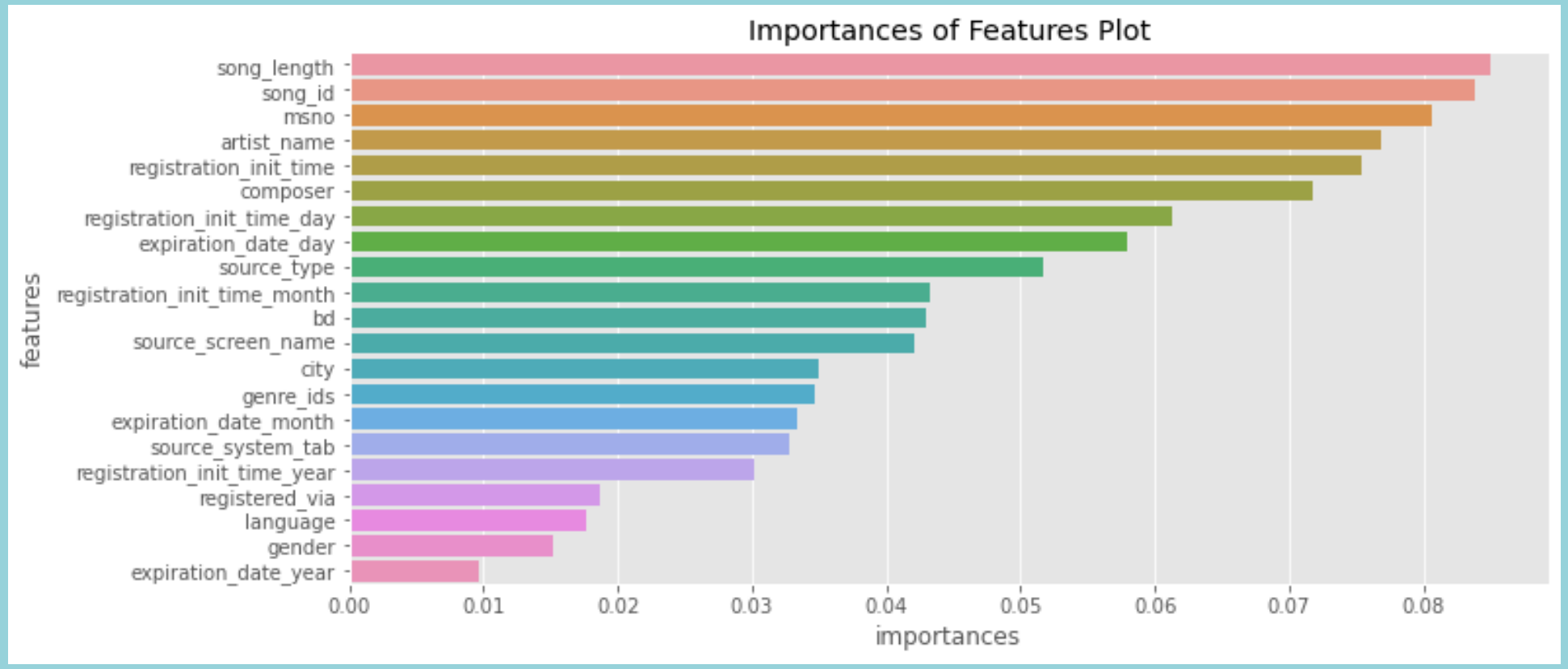
KKBox Dataset

Feature Correlation



- Gender, Language, Source type and Source name had the strongest negative correlations
- While city, artist name and genre had some of the strongest positive correlations

Random Forest Feature Selection



Columns Kept

- Msno
- Song ID
- Source Screen Name – name of tab
- Source Type – entry point for users first listen
- Song Length
- Artist
- Composer
- Age
- Registration month
- Registration day
- Expiration day

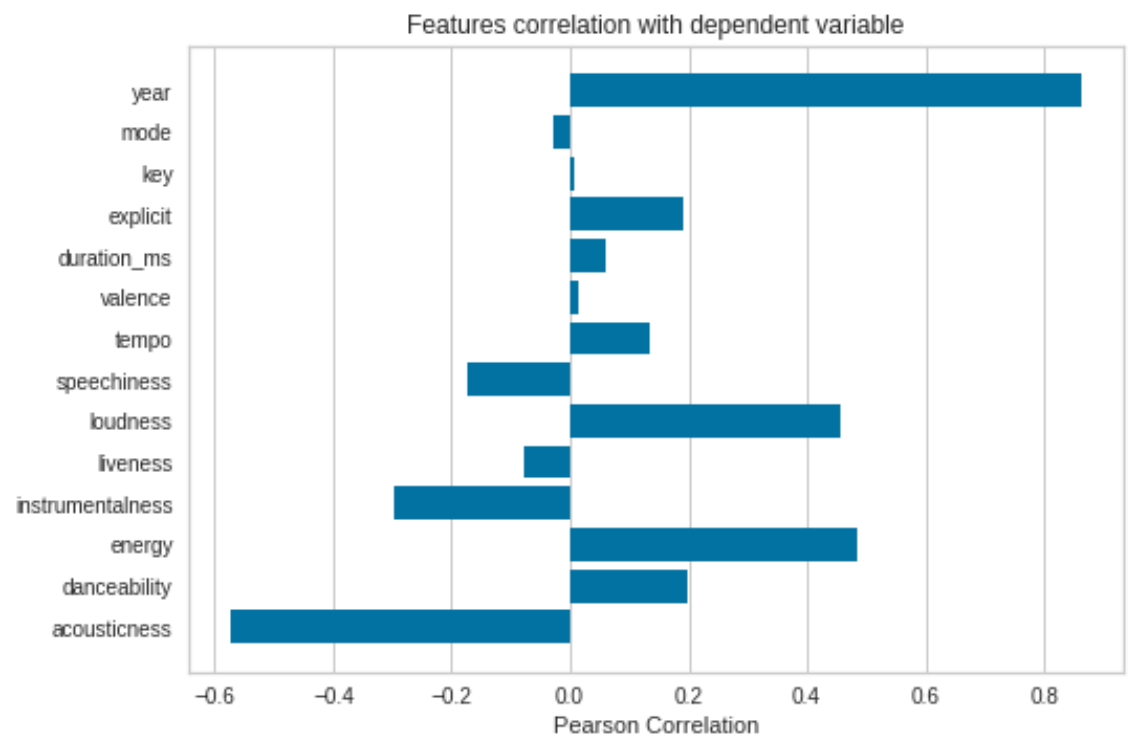
Kkbox XGBoost Results

	Precision	Recall	F1-Score
0	0.63	0.95	0.61
1	0.62	0.65	0.64

Accuracy	0.62

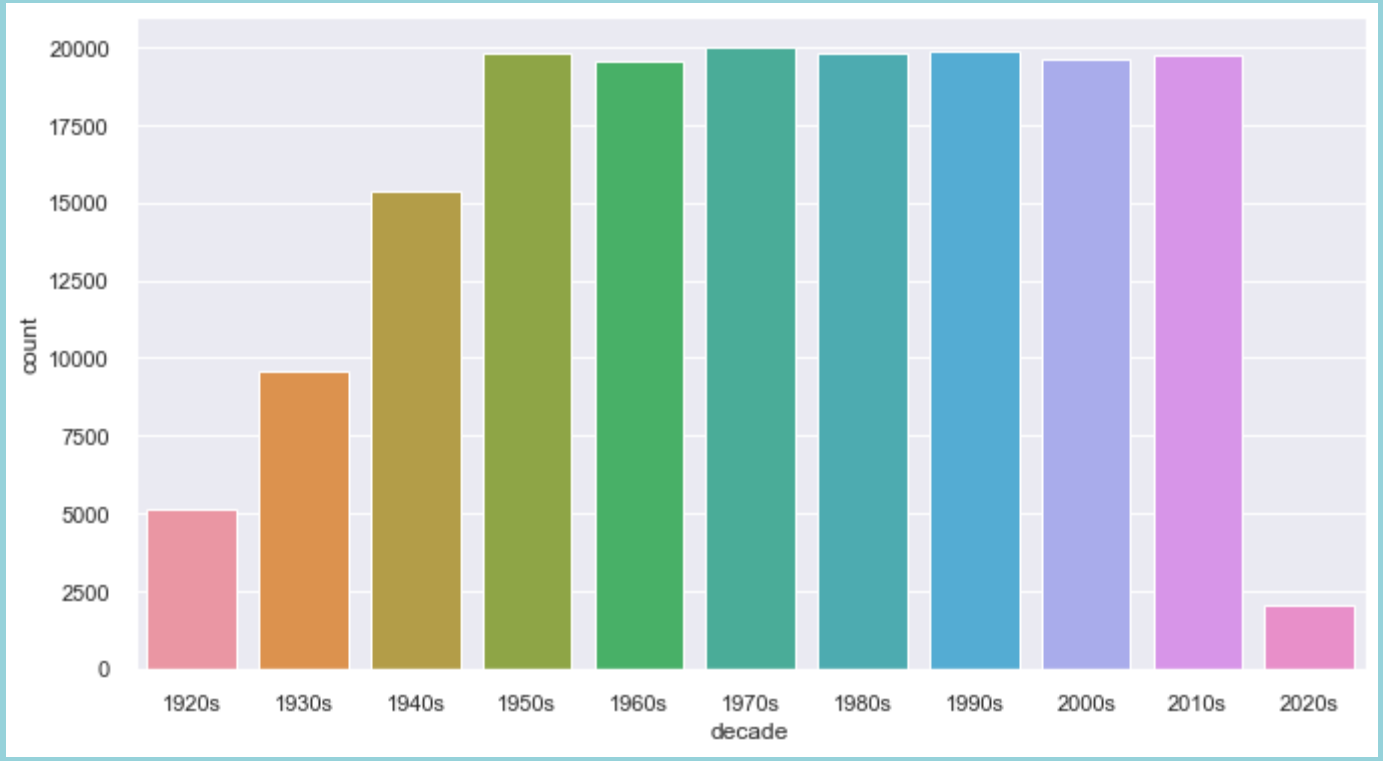
Spotify Dataset

Feature Correlation



- Acousticness, Instrumentalness and Speechiness had the strongest negative correlation
- While Year, Loudness and Energy had the strongest positive correlation

Music Over Time

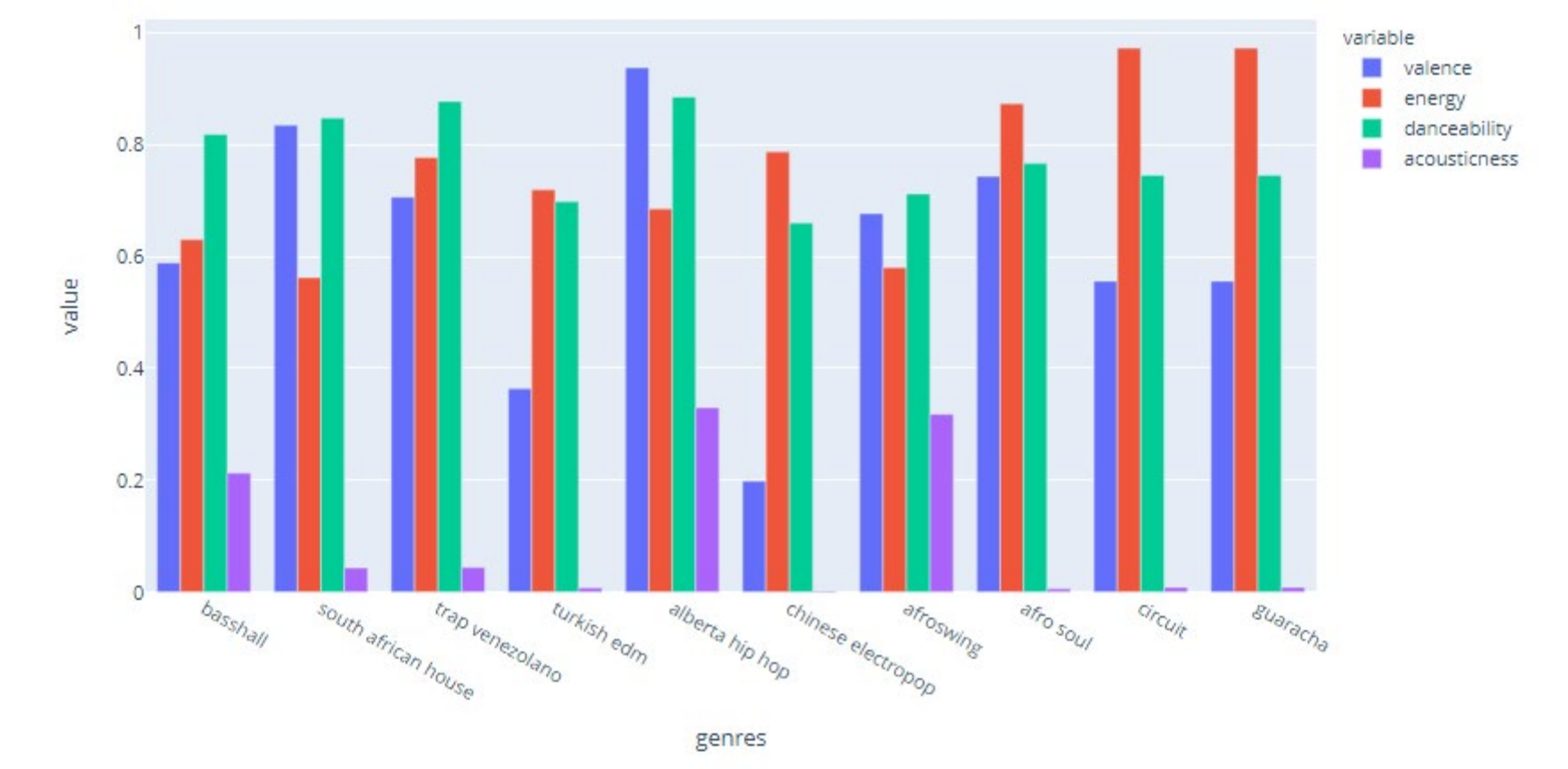


Musical Features Used

- Acousticness
- Danceability
- Energy
- Instrumentalness
- Energy
- Liveness
- Loudness
- Speechiness
- Tempo
- Valence
- Duration ms
- Explicit
- Key
- Mode
- year

Music Over Time

Top 10 Genres



Spotify XGBoost Results

	Precision	Recall	F1-Score
0	0.92	0.92	0.92
1	0.65	0.74	0.69
2	0.56	0.58	0.57
3	0.76	0.62	0.69

Accuracy	0.72

Conclusion

- Based on the accuracy results we can conclude that the Spotify model based on popularity performed better than the KKbox model based on recurring listening events. Given more time and resources the both could be optimized further ,but for now basing our recommendation system on song popularity seems like the way to go.

	Accuracy
Spotify	0.72
KKBox	0.62