Discovering Spotify Wrapped

Credits

People

Performed by

Rita Pereira

Written by

Rita Pereira

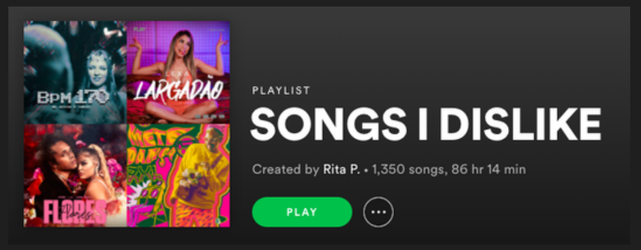
Produced by

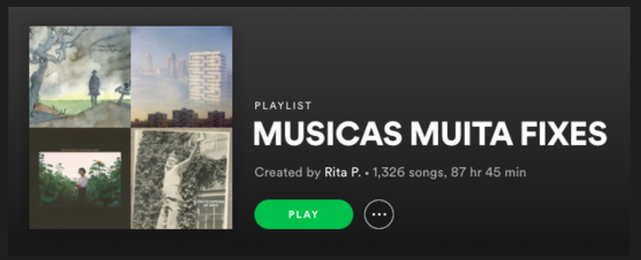
Rita Pereira, Felipa Rocha

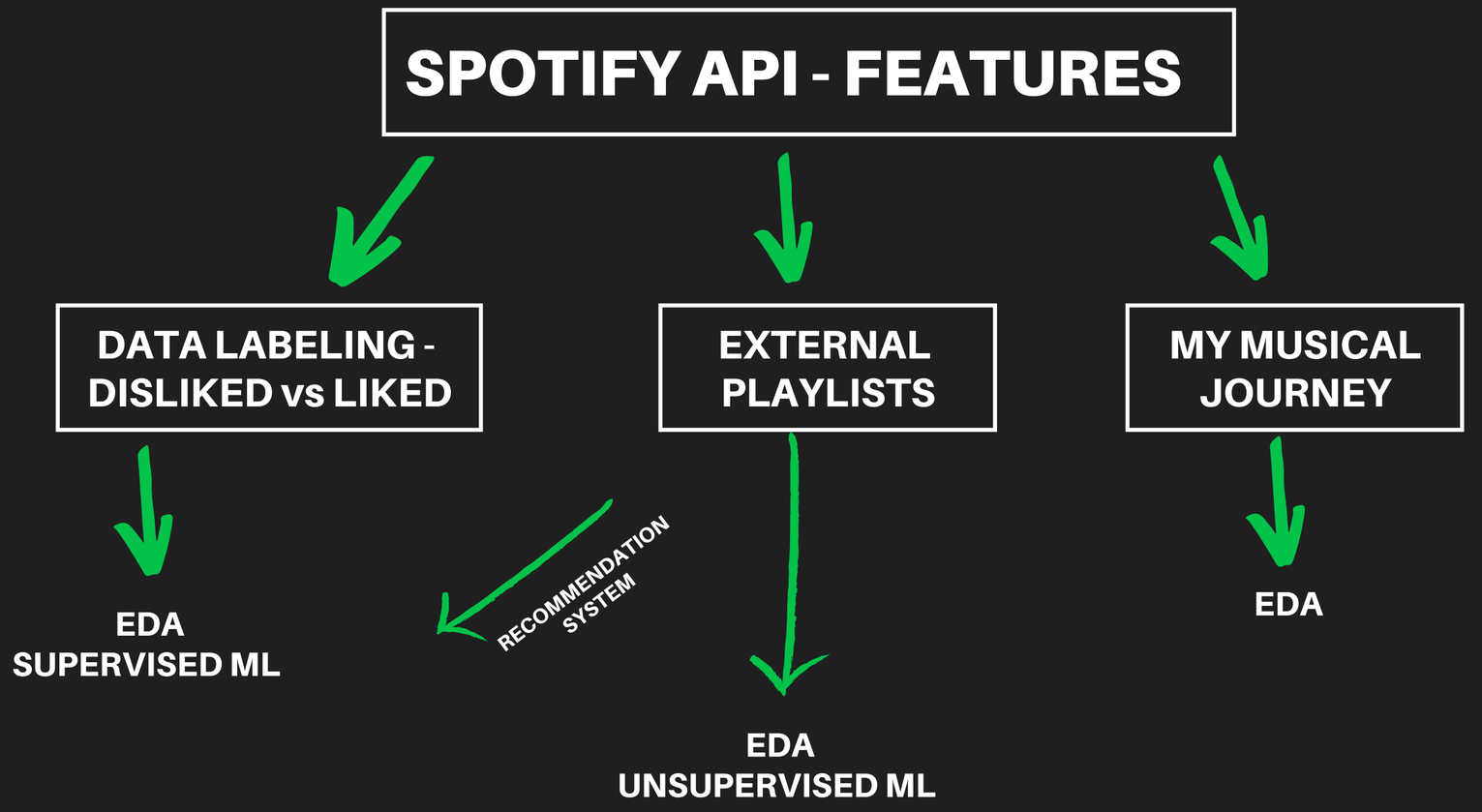
Special Django beats by Pedro Carneiro

Source: Ironhack

Project Workflow



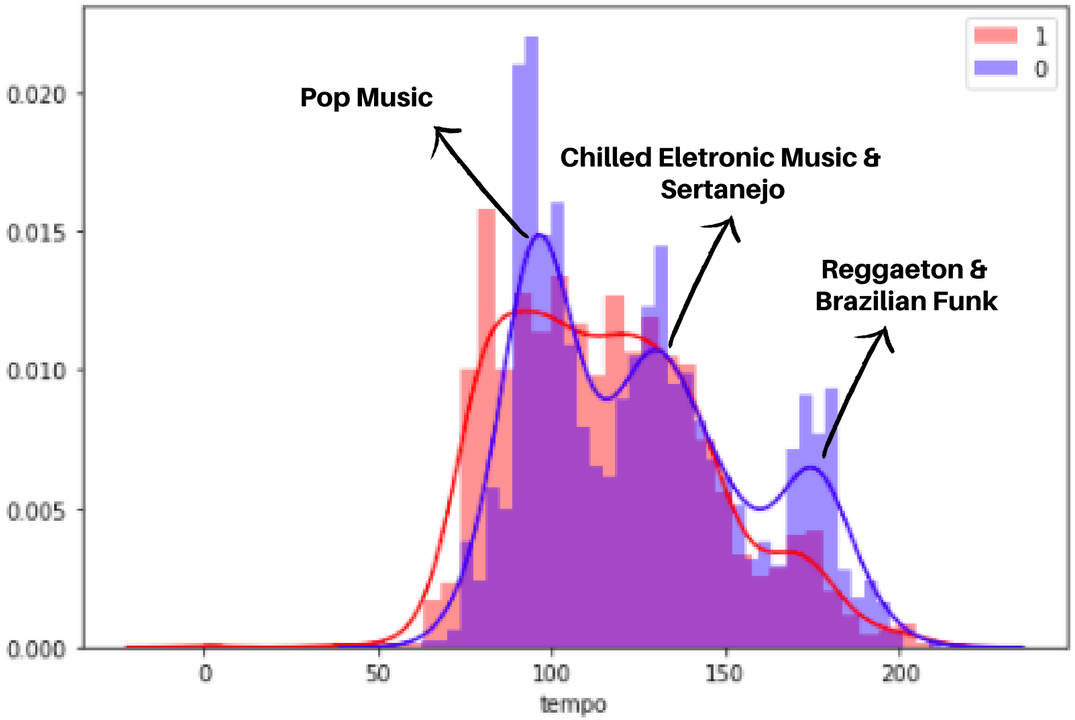




Features & Supervised ML

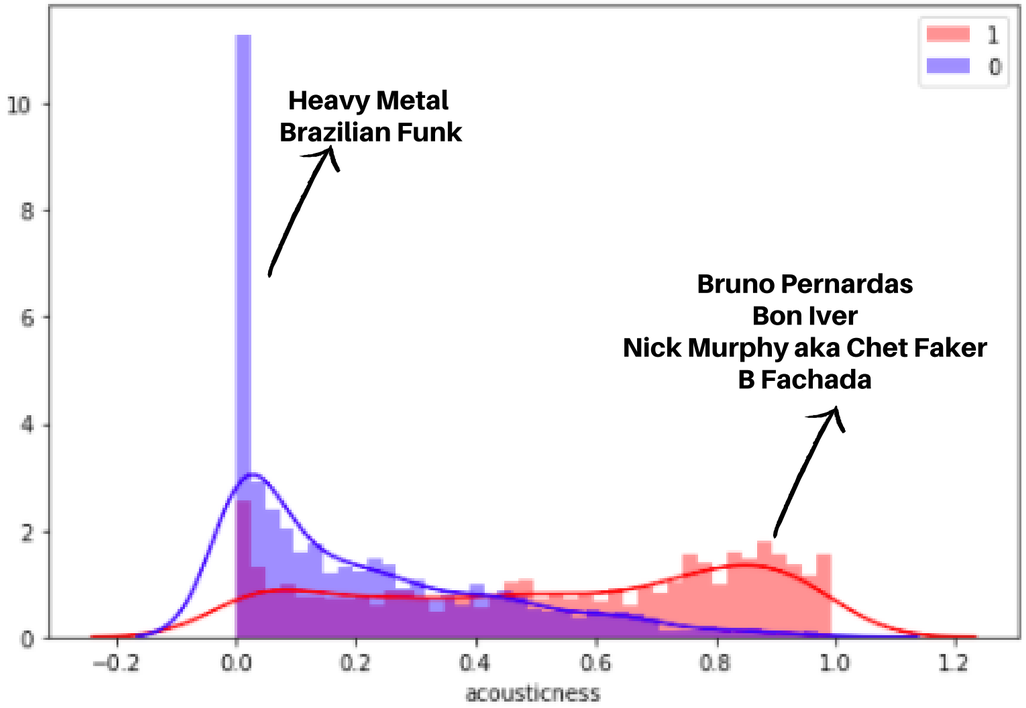
TEMPO.

How many beats per minute (BPM) does each song have?

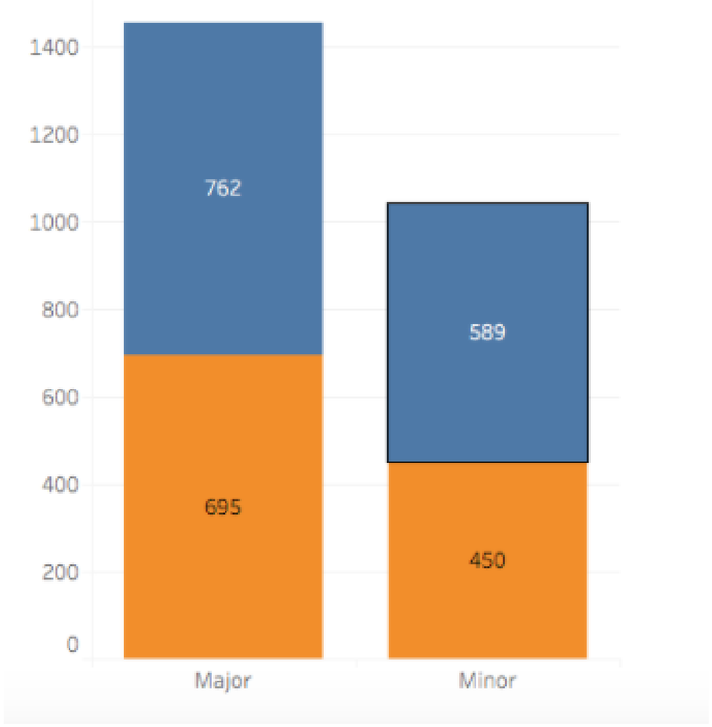


ACOUSTICNESS.

1.0 represents high confidence the track is acoustic.

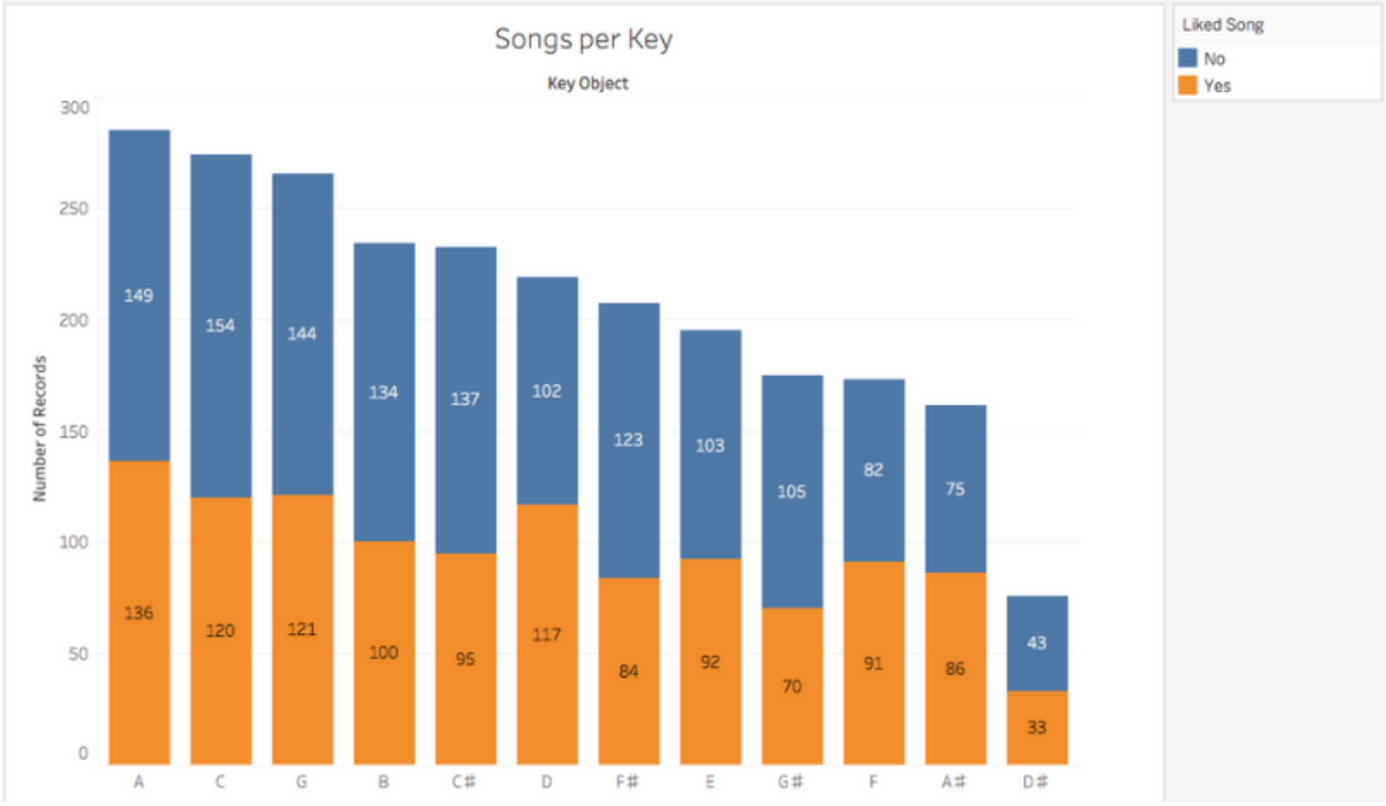


MODE and KEY.



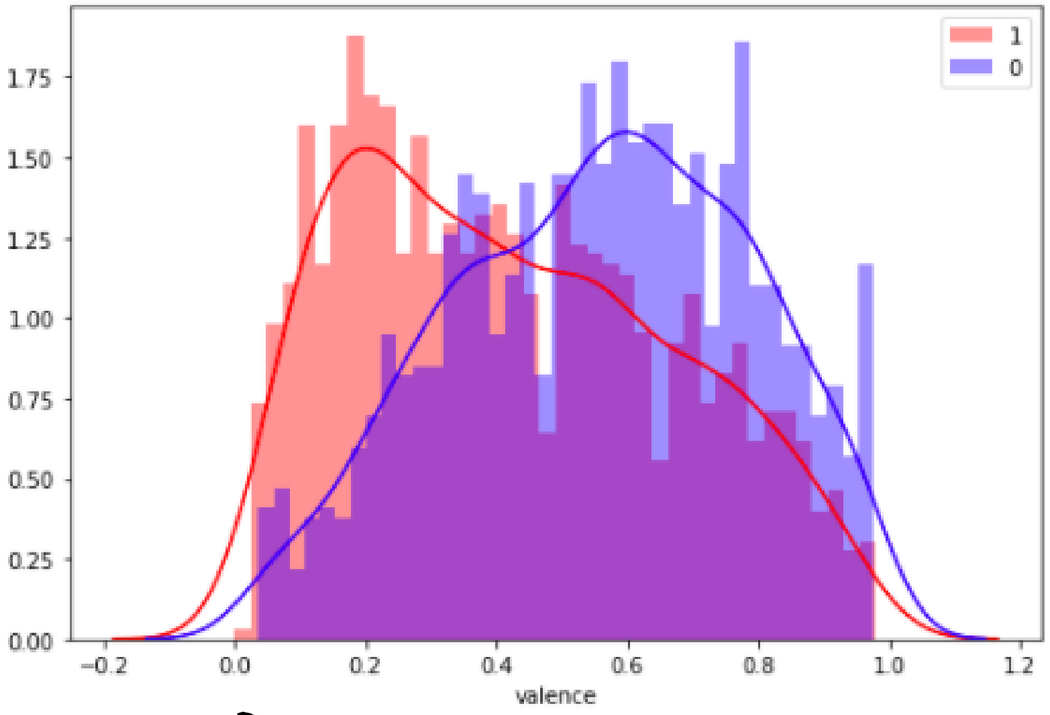
Note one thing from the key description above: "0 = C, 1 = C♯/D ♭, 2 = D"

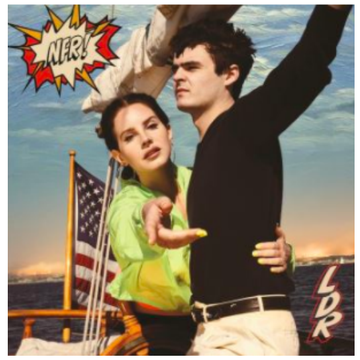
```
data.loc[data['key'] == 0, 'key_object'] = 'C'
data.loc[data['key'] == 1, 'key_object'] = 'C♯'
data.loc[data['key'] == 2, 'key_object'] = 'D'
data.loc[data['key'] == 3, 'key_object'] = 'D♯'
data.loc[data['key'] == 4, 'key_object'] = 'E'#Remember that E and B don't have#
data.loc[data['key'] == 5, 'key_object'] = 'F'
data.loc[data['key'] == 6, 'key_object'] = 'F♯'
data.loc[data['key'] == 7, 'key_object'] = 'G'
data.loc[data['key'] == 8, 'key_object'] = 'G♯'
data.loc[data['key'] == 9, 'key_object'] = 'A'
data.loc[data['key'] == 10, 'key_object'] = 'A♯'
data.loc[data['key'] == 11, 'key_object'] = 'B'
```



VALENCE.

Musical positiveness conveyed by a track.

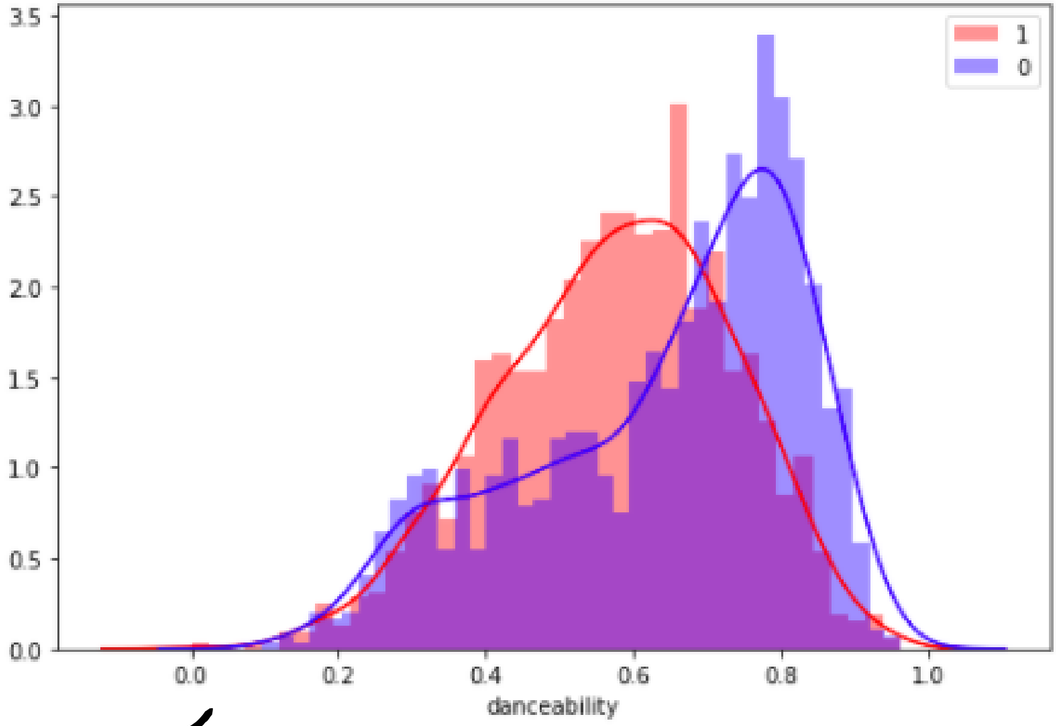




2 Crimes
1 Album

DANCEABILITY.

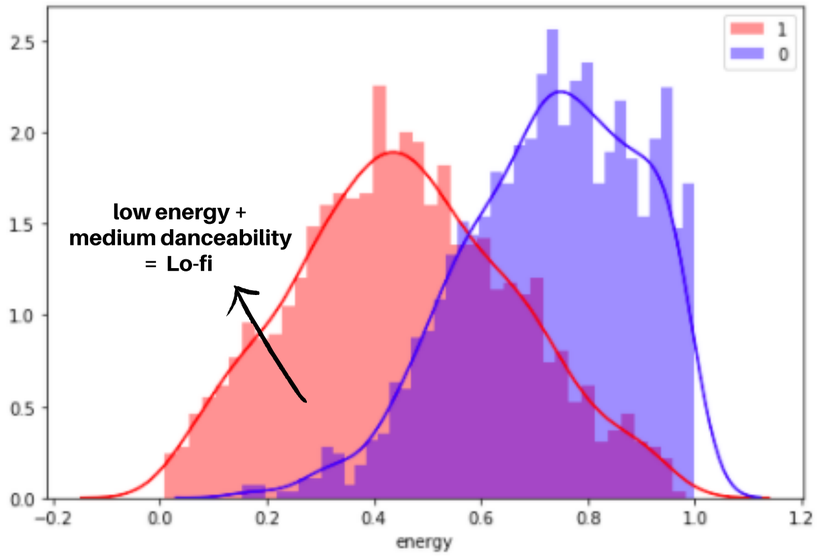
How suitable a track is for dancing?



FYI.
Very Strong Relationship between the two features!!

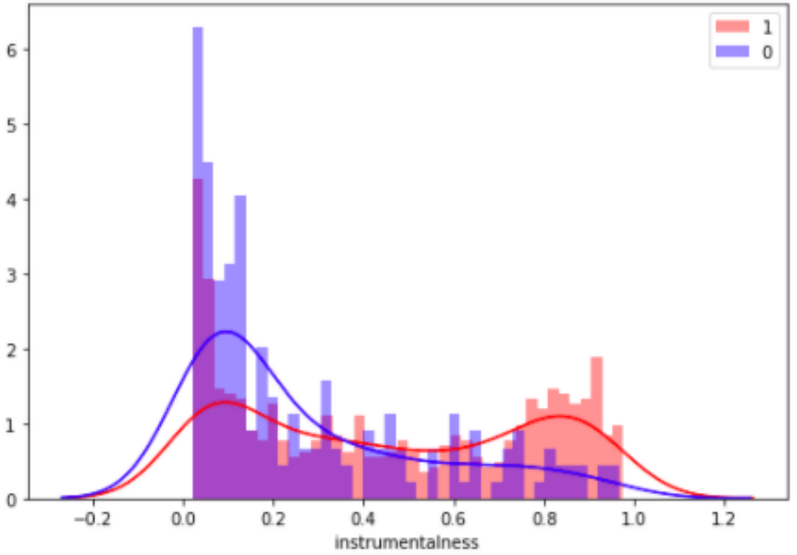
ENERGY.

Measure of intensity and activity.



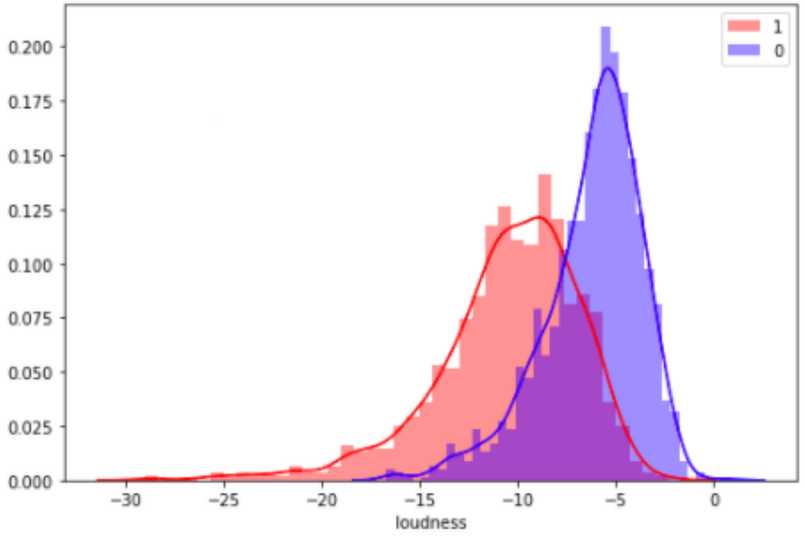
INSTRUMENTALNESS.

Predicts weather a track contains no vocals.



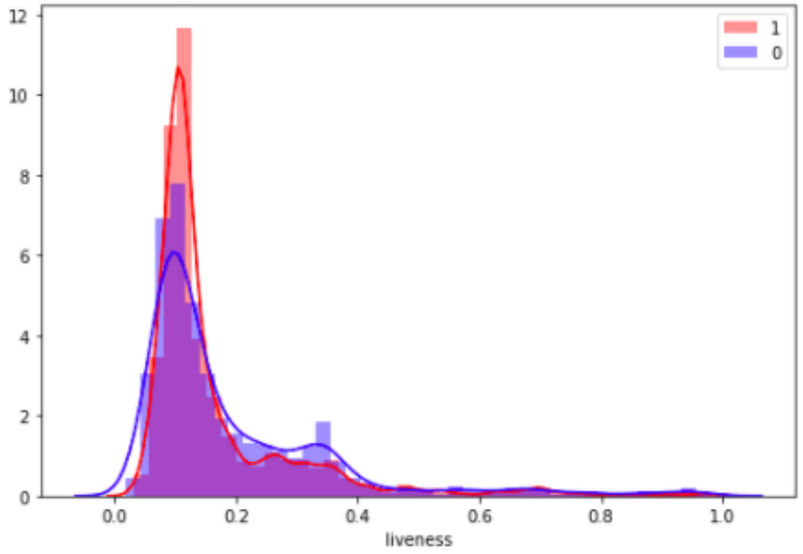
LOUDNESS.

The overall loudness of a track in decibels (dB).



LIVENESS.

Detects the presence of an audience in the recording.



Supervised ML

01 Logistic Regression
Accuracy: 80.0%

02 K-Nearest Neighbors
Accuracy: 80.4 %

03 Decision Tree
Accuracy: 81.9 %

04 Rain Forest
Accuracy: 86.4 %

Supervised ML

Testing Felipe's taste


```
#passing with the Felipe's playlist
pred = clf.predict_proba(felipe_df[features])[:,1]

felipe_df['prediction'] = pred
felipe_df.head()
```

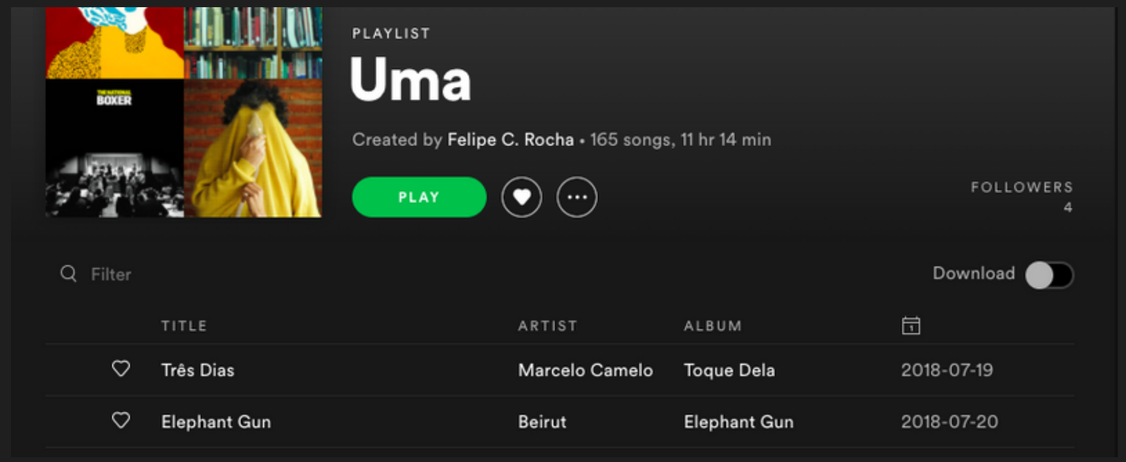
title	main_artist	danceability	energy	key	loudness	mode	acousticness	instrumentalness	liveness	valence	tempo	duration_ms	time_signature	prediction
Três Dias	Marcelo Camelo	0.291	0.319	7	-10.465	1	0.7910	0.46800	0.0692	0.038	103.793	323280	3	0.92
nant Gun	Beirut	0.171	0.626	8	-8.677	1	0.8730	0.02520	0.0681	0.457	180.098	348893	3	0.80
ake pire	The National	0.324	0.776	0	-6.784	1	0.1510	0.91700	0.0728	0.317	101.964	205040	3	0.65
rriso														

04

Rain Forest
Accuracy: 86.4 %

Supervised ML

Testing Felipe's taste




```
#passing with the Felipe's playlist
pred = clf.predict_proba(felipe_df[features])[:,1]

felipe_df['prediction'] = pred
felipe_df.head()
```

title	main_artist	danceability	energy	key	loudness	mode	acousticness	instrumentalness	liveness	valence	tempo	duration_ms	time_signature	prediction
Três Dias	Marcelo Camelo	0.291	0.319	7	-10.465	1	0.7910	0.46800	0.0692	0.038	103.793	323280	3	0.92
nant Gun	Beirut	0.171	0.626	8	-8.677	1	0.8730	0.02520	0.0681	0.457	180.098	348893	3	0.80
ake pire	The National	0.324	0.776	0	-6.784	1	0.1510	0.91700	0.0728	0.317	101.964	205040	3	0.65

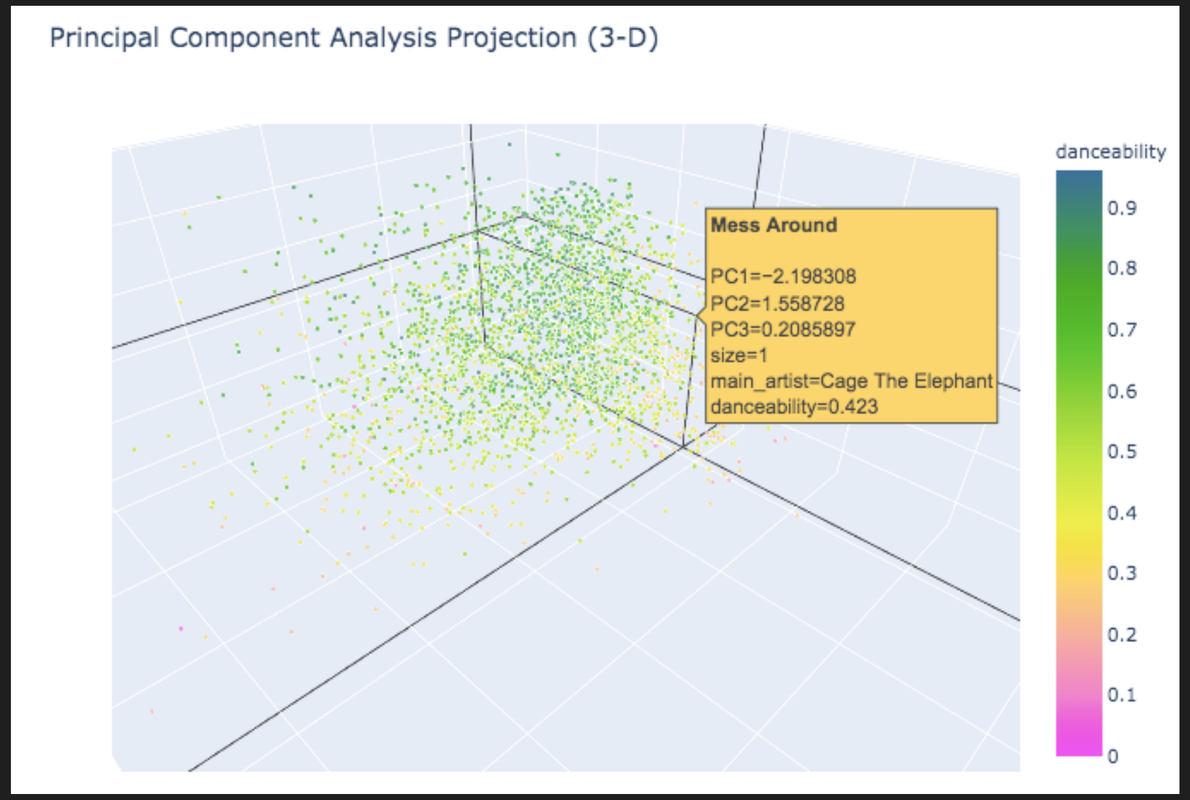
04

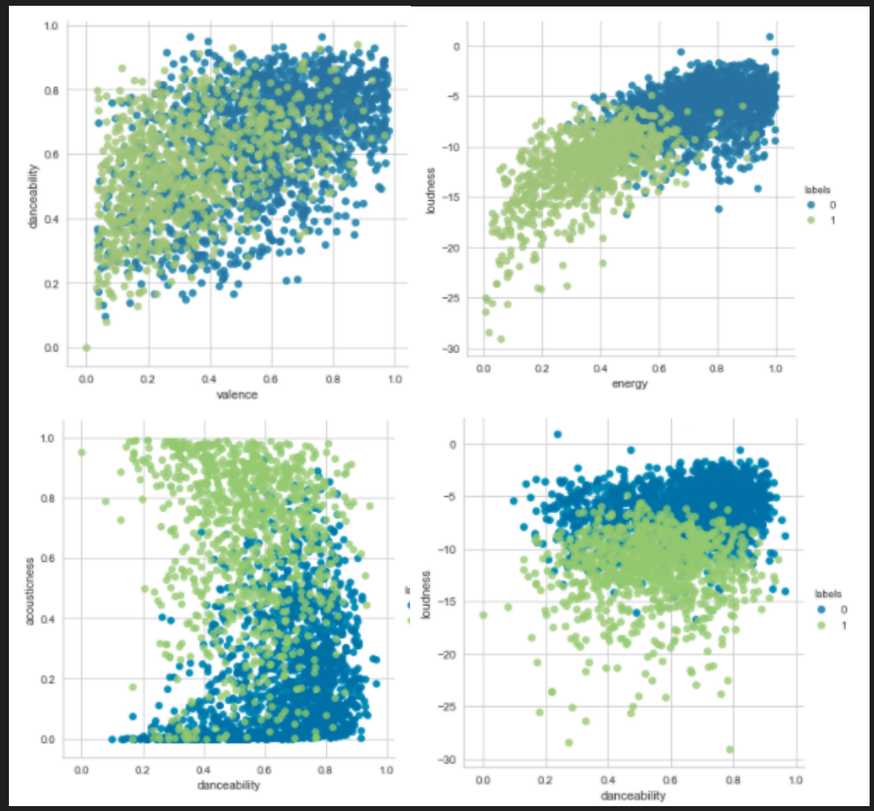
Rain Forest
Accuracy: 86.4 %

APPROVED

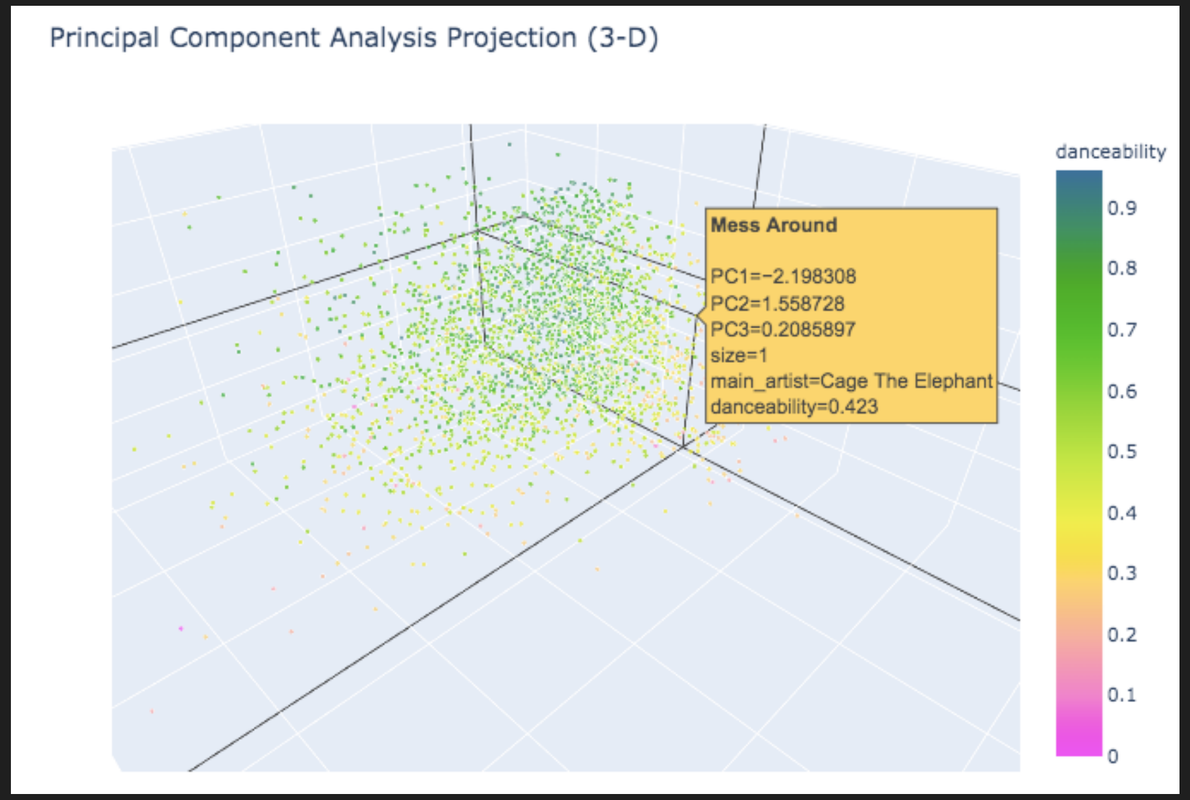
How similar, on average, is it to my taste?
83.5 %

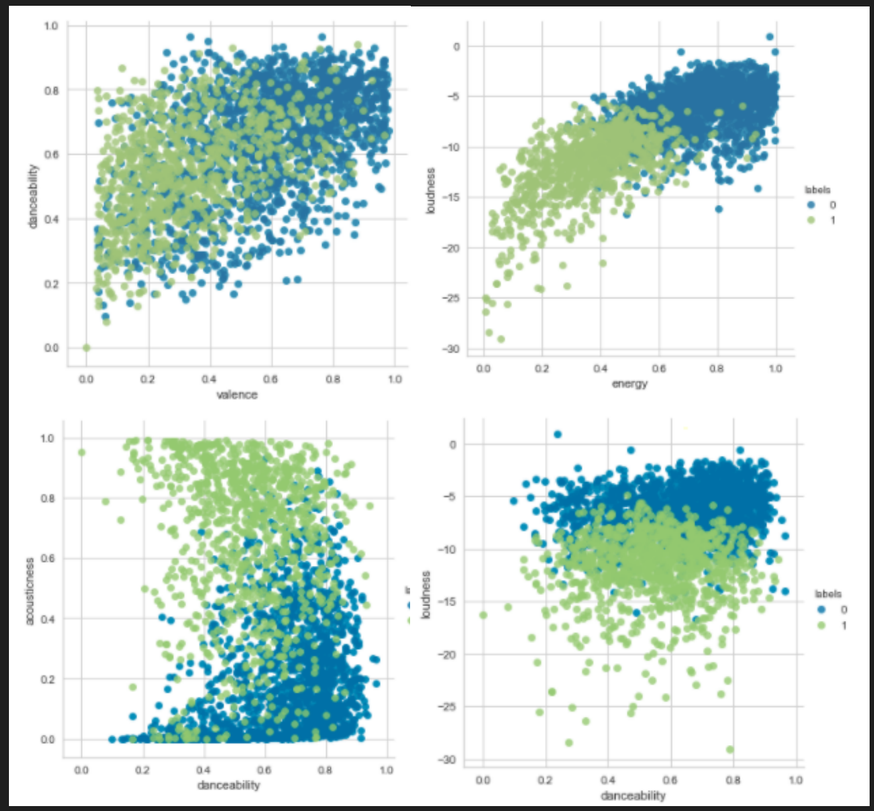
Unsupervised ML & External Playlists





Unsupervised ML & External Playlists





ANY PLAYLIST

+

SONGS DATABASE
> 9000 Song

→

https://aca0d92dcffd.ngrok.io/rita/

DIFFICULTIES

1- GETTING DISTRACTED DOING PLAYLISTS INSTEAD OF THE ACTUAL PROJECT

2- PRETENDING TO BE A WEB DEVELOPER

FURTHER WORK

1- CONNECT ALL MY FRIENDS DISCOVERY WEEKLY PLAYLISTS

2- IMPROVE THE DJANGO APP

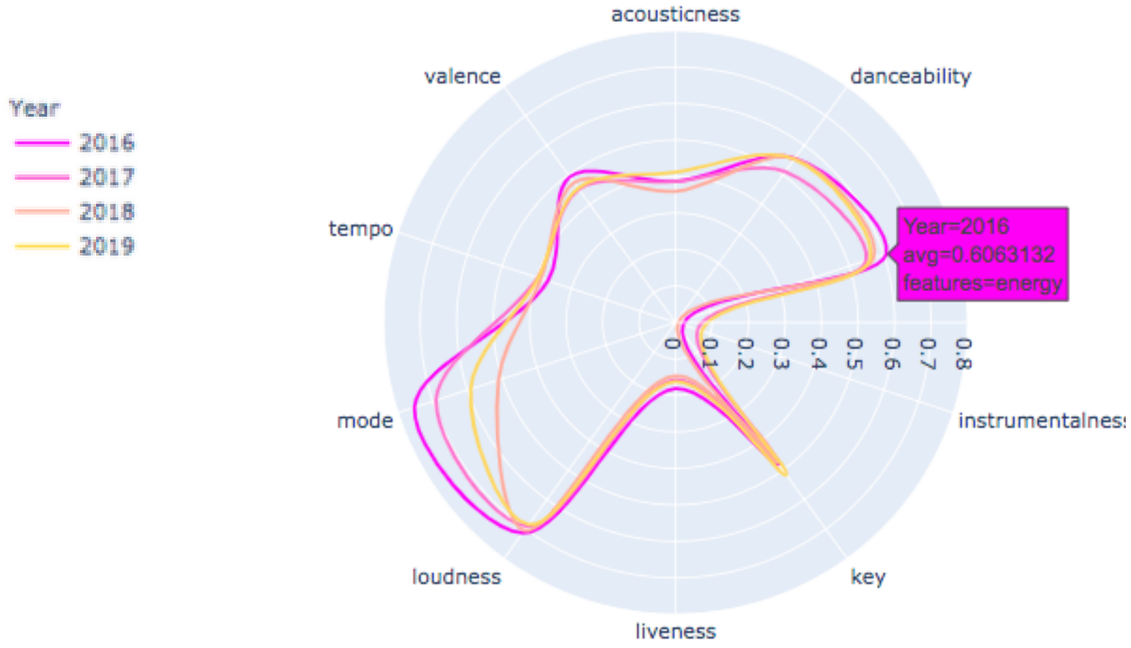
3- USE NLP TO IMPROVE THE MODEL

...

MY MUSICAL JOURNEY

WITH SPOTIFY WRAPPED.

Mean Values of Each Track Features



Principal Component Analysis Projection (3-D)