PROJECT
PROPOSAL

billboard top 100 chart predictions model

mid project

BY:
Zeina Mousa

context

This dataset is designed to predict the future performance of songs on the charts, including their future rankings, longevity, and peak positions. It also aims to identify the factors influencing each track's chart success.

DATA COLLECTION SOURCES

- Spotify DevTools
- Spotify API
- musicbrainz
- github repository (billboard top 100)

ATTRIBUTE OVERVIEW

chart week
the chart's date (weekly)

current week
the song's current position

title
the track's title

performer(s)
the song's performer(s)

last week
Previous week's position

peak_pos
the song's peak position

wks_on_chart
weeks the song spent in the top 100

genre(s)
the artist's typical genre(s)

duration-ms
the song's duration in milliseconds

tempo
The overall estimated tempo of a track in beats per minute (BPM)

loudness
The overall loudness of a track in decibels (dB). Values typically range between -60 and 0 db.

energy
Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity.

danceability
how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity.

Data overview

	current_week	peak_pos	wks_on_chart	duration-ms	tempo	loudness	energy	danceability
count	27700.000000	27700.000000	27700.000000	2.770000e+04	27699.000000	27699.000000	27699.000000	27699.000000
mean	50.499964	33.743755	12.315993	2.069252e+05	122.978162	-6.868097	0.632008	0.656826
std	28.866631	28.010730	11.774950	1.409496e+05	29.210766	2.606790	0.161173	0.144776
min	1.000000	1.000000	1.000000	2.890000e+04	53.376000	-26.749000	0.031600	0.128000
25%	25.750000	9.000000	3.000000	1.694180e+05	99.894000	-7.984000	0.530000	0.555000
50%	50.500000	27.000000	9.000000	1.935060e+05	121.987000	-6.335000	0.643000	0.661000
75%	75.250000	55.000000	17.000000	2.215200e+05	142.868000	-5.092000	0.754000	0.765000
max	100.000000	100.000000	91.000000	3.614013e+06	215.338000	-1.896000	0.991000	0.971000

	chart_week	title	performer	last_week	genre(s)
count	27700	27700	27700	25884	16173
unique	277	3447	1796	101	354
top	1/4/2020	Heat Waves	Morgan Wallen	-	country
freq	100	91	712	2472	3232

Data cleaning required

ASSIGNING COLUMNS TO THEIR APPROPRIATE DATA TYPE

- chart week : date
- last week: int64

FILLING NULL VALUES WITH THE APPROPRIATE CONTENT

- genre(s): filling null values from another website

#	Column	Non-Null Count	Dtype
0	chart_week	27700 non-null	object
1	current_week	27700 non-null	int64
2	title	27700 non-null	object
3	performer	27700 non-null	object
4	last_week	25884 non-null	object
5	peak_pos	27700 non-null	int64
6	wks_on_chart	27700 non-null	int64
7	genre(s)	16173 non-null	object
8	explicit	27700 non-null	bool
9	duration-ms	27700 non-null	float64
10	tempo	27699 non-null	float64
11	loudness	27699 non-null	float64
12	energy	27699 non-null	float64
13	danceability	27699 non-null	float64

The data collected will be analyzed to consider each factor in the track to see which have the most influence regarding the track's position in the chart. With that in mind and the appropriate algorithm chosen the system should be able to accurately predict next week's chart in regards to each track's position (with the consideration of new releases)