Top Hit Consultants

A music data journey using Python

Table of contents

The problem

We are Top Hit, a consulting company for independent artists. We are helping a new Mexican artist to launch their next hit in **Mexico**, which is planned to be released in last quarter 2020 in the **Spotify** platform, to remain in the **Top 50 Chart** throughout the next year.

The question we are trying to solve is:

What features should this artist’s song have to be successful throughout the next year?

Our data

- Spotify Mexico Top 50 Daily Chart (2017 – 2019)
- Spotify genre dataset
- Spotify API song data

Relevant features:

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

Energy: A measure from 0.0 to 1.0 that represents a perceptual measure of intensity and activity

Tempo: The rate of speed of a musical piece or passage, often indicated by an exact metronome marking (bpm)

Total appearances: Total appearances of a given song in the Top 50 Daily Chart data set.

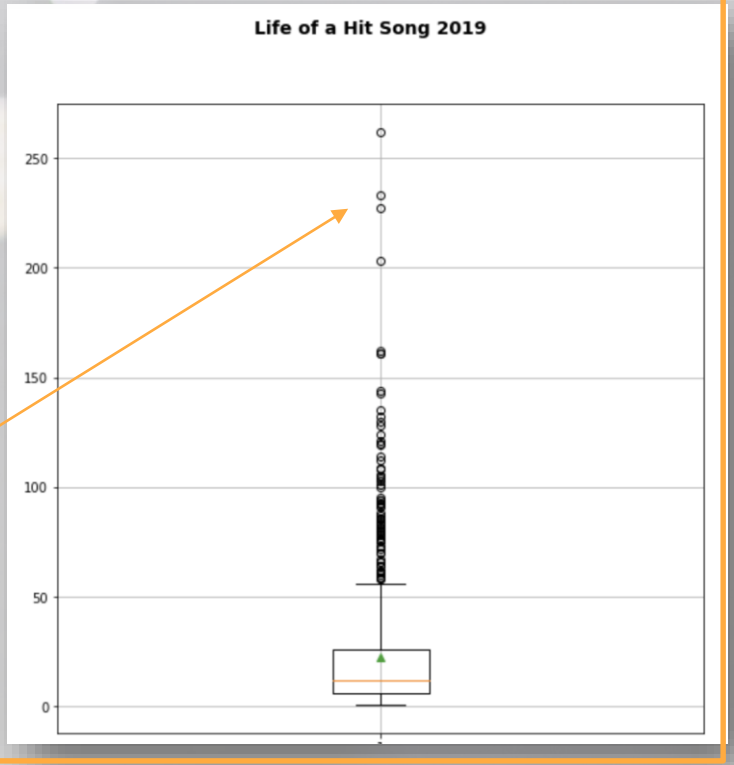
Unique appearances: This subset of the Total appearances counting each song in the data set only once.

Song highlights

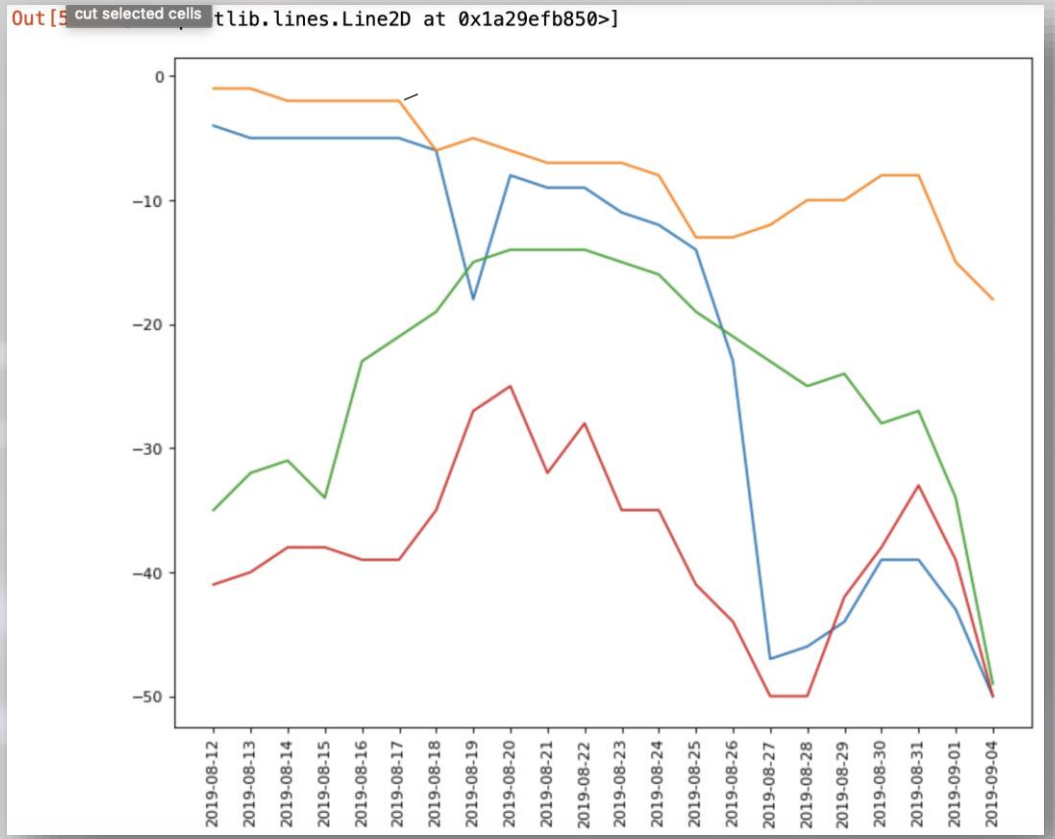
- Merge of Top Hit Databases from 2017 to 2019 with song's features and song's genres
- The the median life of a song in Top 50 is: 12.0 days
- … but outliers start at 56 days of presence in the chart

Out[51]:

	Artist	Track Name	Number of days in Top 50 Chart
2132	LP	Lost on You	370
1099	Ed Maverick	Fuentes de Ortiz	357
1700	Jack Stauber	Buttercup	256
3829	Tones And I	Dance Monkey	244
639	Carin Leon	Me La Avente	233



Song highlights



Song highlights

```
In [69]: season1=analysis[(analysis['Date']>'2017-05-10') & (analysis['Date']<'2017-05-17')]
season2=analysis[(analysis['Date']>'2018-05-10') & (analysis['Date']<'2018-05-17')]
season3=analysis[(analysis['Date']>'2019-05-10') & (analysis['Date']<'2019-05-17')]
seasonu=pd.concat([season1, season2,season3])
seasonu=seasonu.sort_values(by='Position', ascending=True)
seasonu = seasonu[(seasonu['Artist'] == "Denise De Kalafe")]
seasonu
```

Out[69]:

	Position	Track Name	Artist	URL_x	Date	id_x	
6651	2	Señora... Señora... Señora	Denise De Kalafe	https://open.spotify.com/track/1gpnjVCS2P9BPug...	2017-05-14	1gpnjVCS2P9BPug3FvHpgW	ht
6551	2	Señora... Señora...	Denise De	https://open.spotify.com/track/1gpnjVCS2P9BPug...	2017-05-12	1gpnjVCS2P9BPug3FvHpgW	ht
6601	2	Señora... Señora... Señora	Denise De Kalafe	https://open.spotify.com/track/1gpnjVCS2P9BPug...	2017-05-13	1gpnjVCS2P9BPug3FvHpgW	ht
6501	2	Señora... Señora... Señora	Denise De Kalafe	https://open.spotify.com/track/1gpnjVCS2P9BPug...	2017-05-11	1gpnjVCS2P9BPug3FvHpgW	ht
24752	3	Señora... Señora... Señora	Denise De Kalafe	https://open.spotify.com/track/1gpnjVCS2P9BPug...	2018-05-13	1gpnjVCS2P9BPug3FvHpgW	
24702	3	Señora... Señora... Señora	Denise De Kalafe	https://open.spotify.com/track/1gpnjVCS2P9BPug...	2018-05-12	1gpnjVCS2P9BPug3FvHpgW	
24652	3	Señora... Señora... Señora	Denise De Kalafe	https://open.spotify.com/track/1gpnjVCS2P9BPug...	2018-05-11	1gpnjVCS2P9BPug3FvHpgW	
6754	5	Señora... Señora... Señora	Denise De Kalafe	https://open.spotify.com/track/1gpnjVCS2P9BPug...	2017-05-16	1gpnjVCS2P9BPug3FvHpgW	
6704	5	Señora... Señora... Señora	Denise De Kalafe	https://open.spotify.com/track/1gpnjVCS2P9BPug...	2017-05-15	1gpnjVCS2P9BPug3FvHpgW	
24804	5	Señora... Señora... Señora	Denise De Kalafe	https://open.spotify.com/track/1gpnjVCS2P9BPug...	2018-05-14	1gpnjVCS2P9BPug3FvHpgW	

click to scroll output; double click to hide

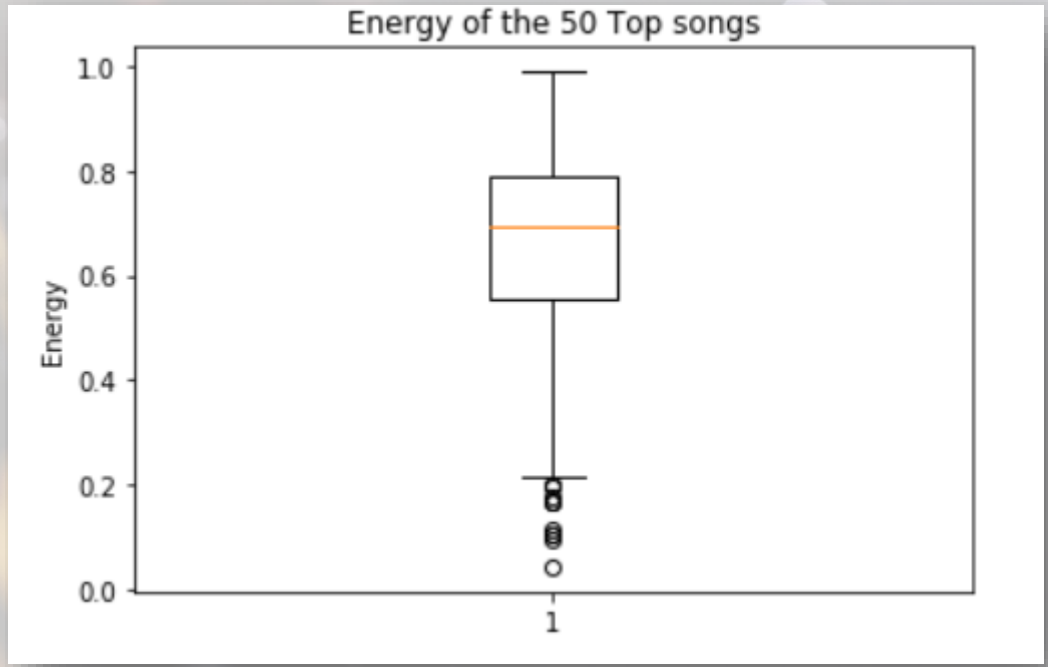
Energy analysis

Hypothesis: If a song has a higher level of energy, then the song will appear more times in the Top 50 throughout the next year.

Track Name	Number of times in the chart TOP 50	Energy
11 Minutes (with Halsey feat. Travis Barker)	10	0.852
13 Beaches	8	0.402
1950	14	0.535
1999	6	0.730
24/7	4	0.740
...	...	...
the light is coming (feat. Nicki Minaj)	8	0.544
the remedy for a broken heart (why am I so in love)	5	0.304
when the party's over	23	0.104
wish you were gay	15	0.351
¿Qué Tiene?	84	0.534

The **correlation** between energy and Number of times in the chart TOP 50 factors is **0.1**

Energy boxplot



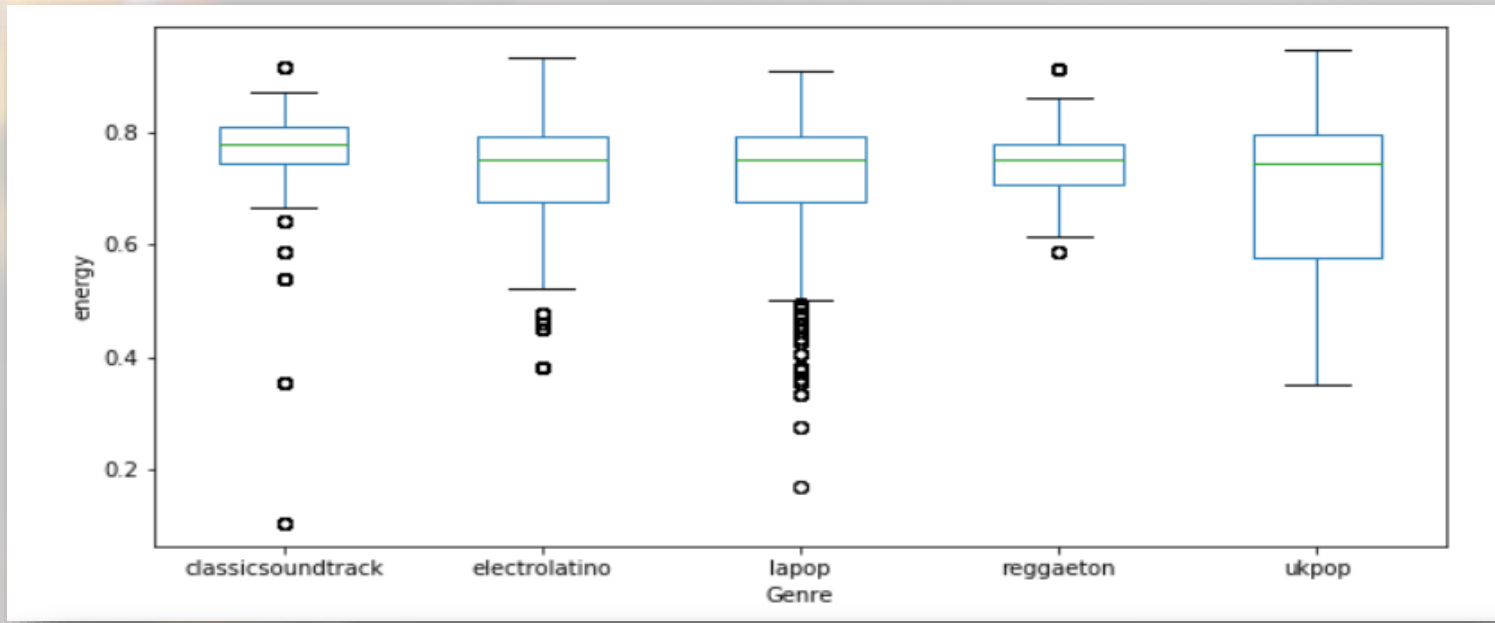
Removing the energy outliers

```
energy_wo=merge5_df[(merge5_df>lower_bound_energy).all(1)]
energy_wo
```

Track Name	Number of times in the chart TOP 50	Energy
11 Minutes (with Halsey feat. Travis Barker)	10	0.852000
13 Beaches	8	0.402000
1950	14	0.535000
1999	6	0.730000
24/7	4	0.740000
...	...	...
rockstar	143	0.523448
the light is coming (feat. Nicki Minaj)	8	0.544000
the remedy for a broken heart (why am I so in love)	5	0.304000
wish you were gay	15	0.351000
¿Qué Tiene?	84	0.534000

The correlation between energy and the number of times in the chart Top 50 without the outliers is **0.09**.

Boxplot of the energy of the TOP 5 genres




```
stats.f_oneway(g1,g2,g3,g4,g5)
```

```
F_onewayResult(statistic=305.6158719189037, pvalue=3.1094655330479858e-257)
```

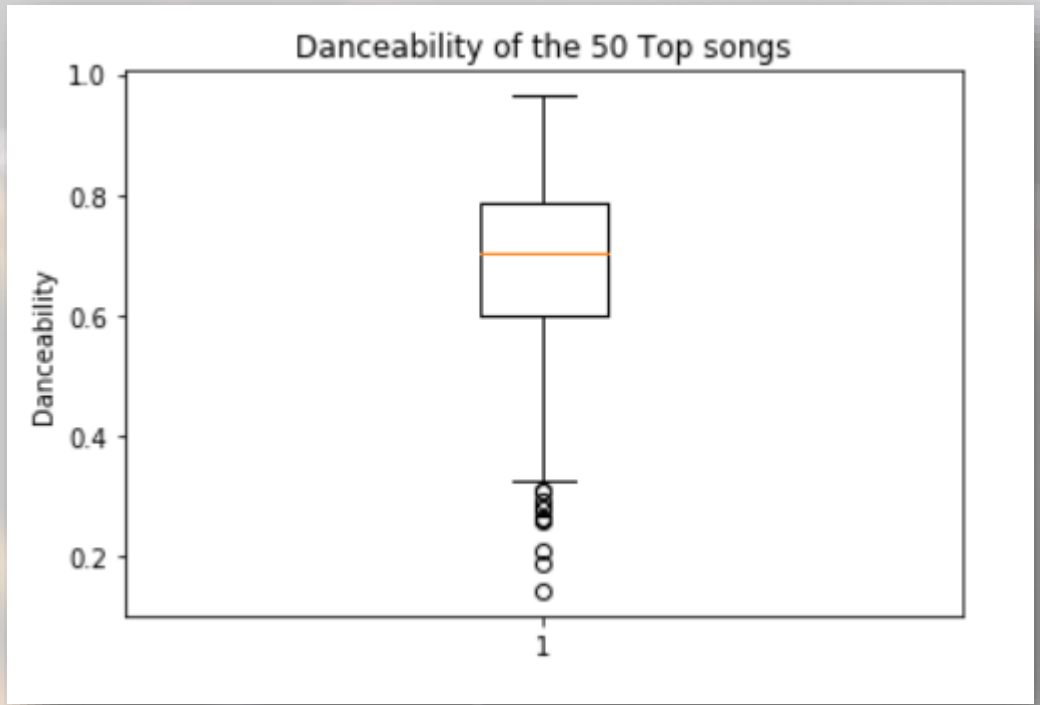
Danceability analysis

Hypothesis: If a song has a higher level of energy, then the song will appear more times in the Top 50 throughout the next year.

Track Name	Number of times in the chart TOP 50	Danceability
11 Minutes (with Halsey feat. Travis Barker)	10	0.464
13 Beaches	8	0.415
1950	14	0.600
1999	6	0.866
24/7	4	0.615
...	...	...
the light is coming (feat. Nicki Minaj)	8	0.879
the remedy for a broken heart (why am I so in love)	5	0.699
when the party's over	23	0.498
wish you were gay	15	0.853
¿Qué Tiene?	84	0.708

The correlation between danceability and number of times in the chart Top 50 is **0.19**

Danceability boxplot



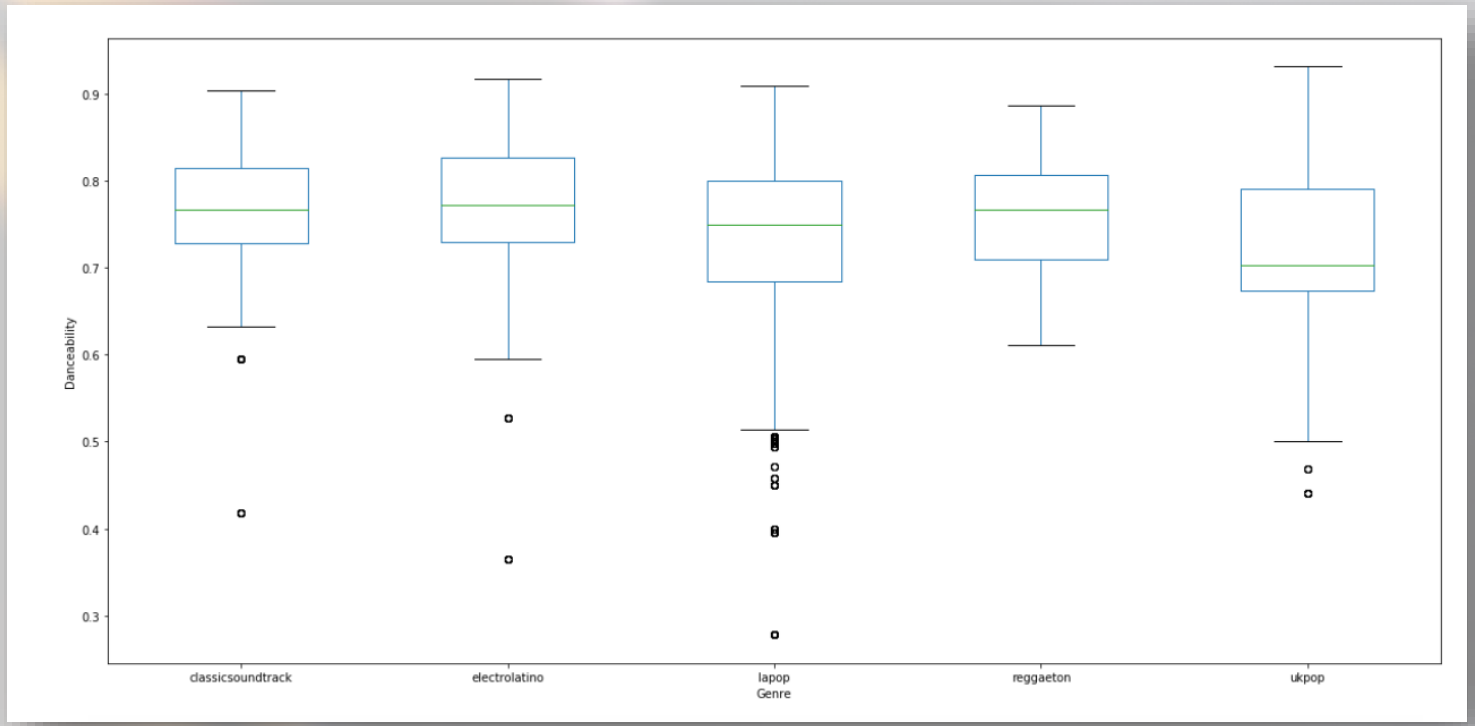
Removing the danceability outliers

```
dance_wo=merge_df[(merge_df>=lower_bound_dance).all(1)]
dance_wo
```

Track Name	Number of times in the chart TOP 50	Danceability
11 Minutes (with Halsey feat. Travis Barker)	10	0.464
13 Beaches	8	0.415
1950	14	0.600
1999	6	0.866
24/7	4	0.615
...	...	...
the light is coming (feat. Nicki Minaj)	8	0.879
the remedy for a broken heart (why am I so in love)	5	0.699
when the party's over	23	0.498
wish you were gay	15	0.853
¿Qué Tiene?	84	0.708

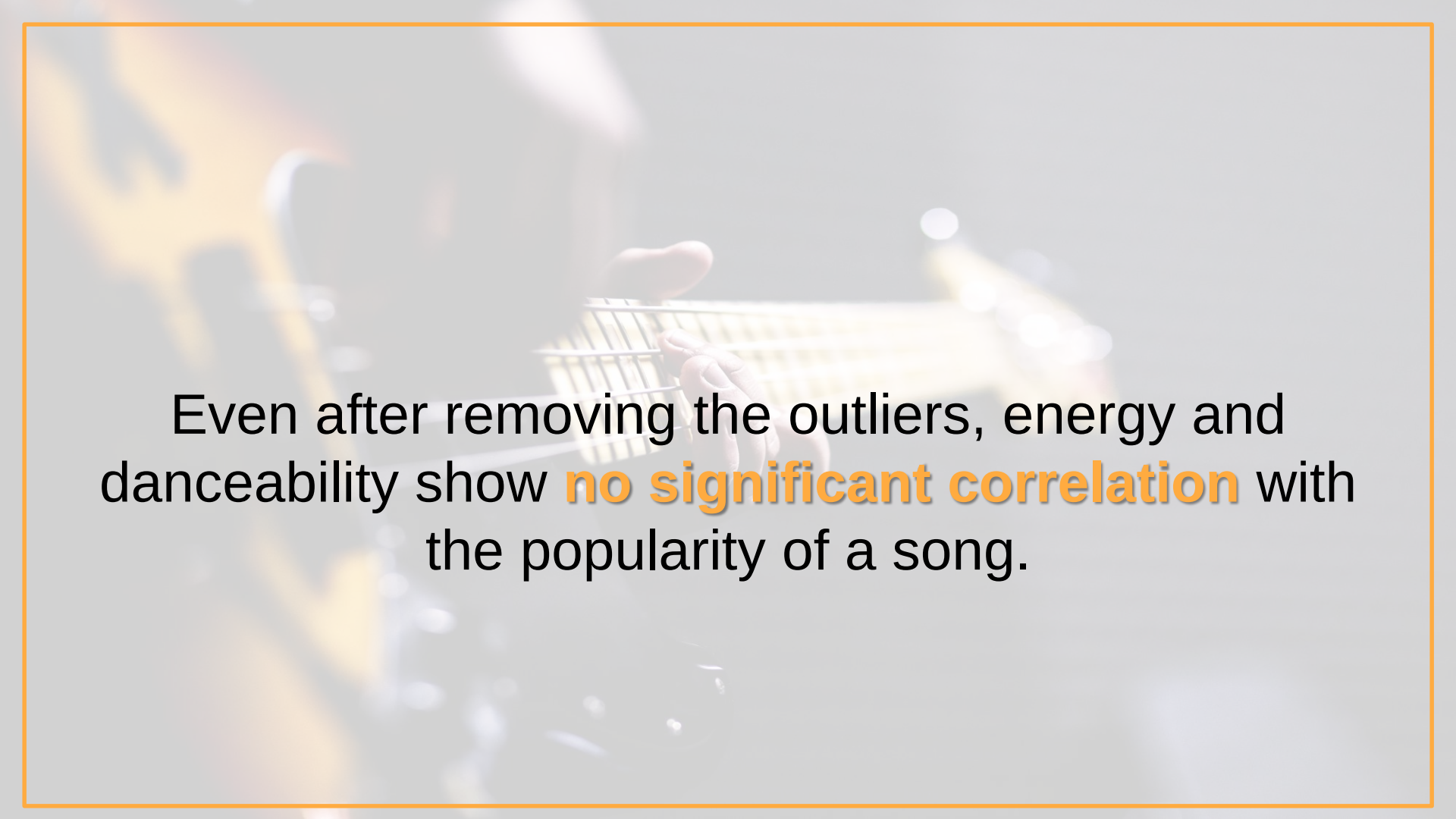
The correlation between danceability and the number of times in the chart Top 50 without the outliers is **0.19**.

Boxplot of the danceability of the TOP 5 genres



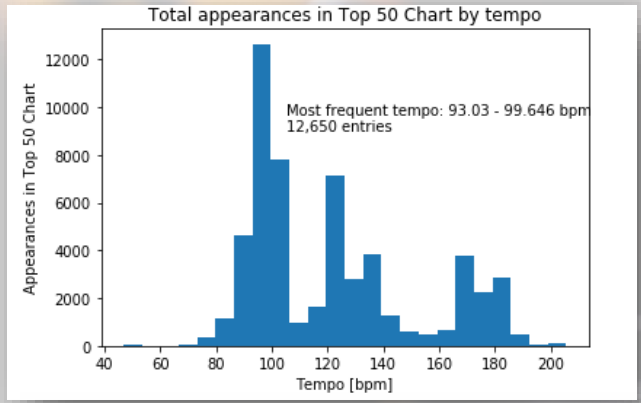

```
stats.f_oneway(g1,g2,g3,g4,g5)
```

```
F_onewayResult(statistic=305.6158719189037, pvalue=3.1094655330479858e-257)
```



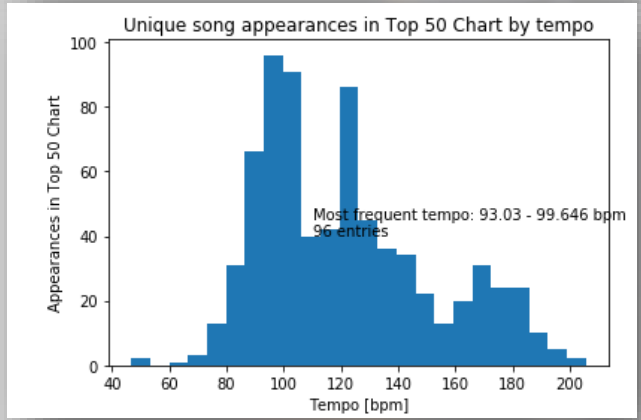
Even after removing the outliers, energy and danceability show **no significant correlation** with the popularity of a song.

Music Tempo Analysis



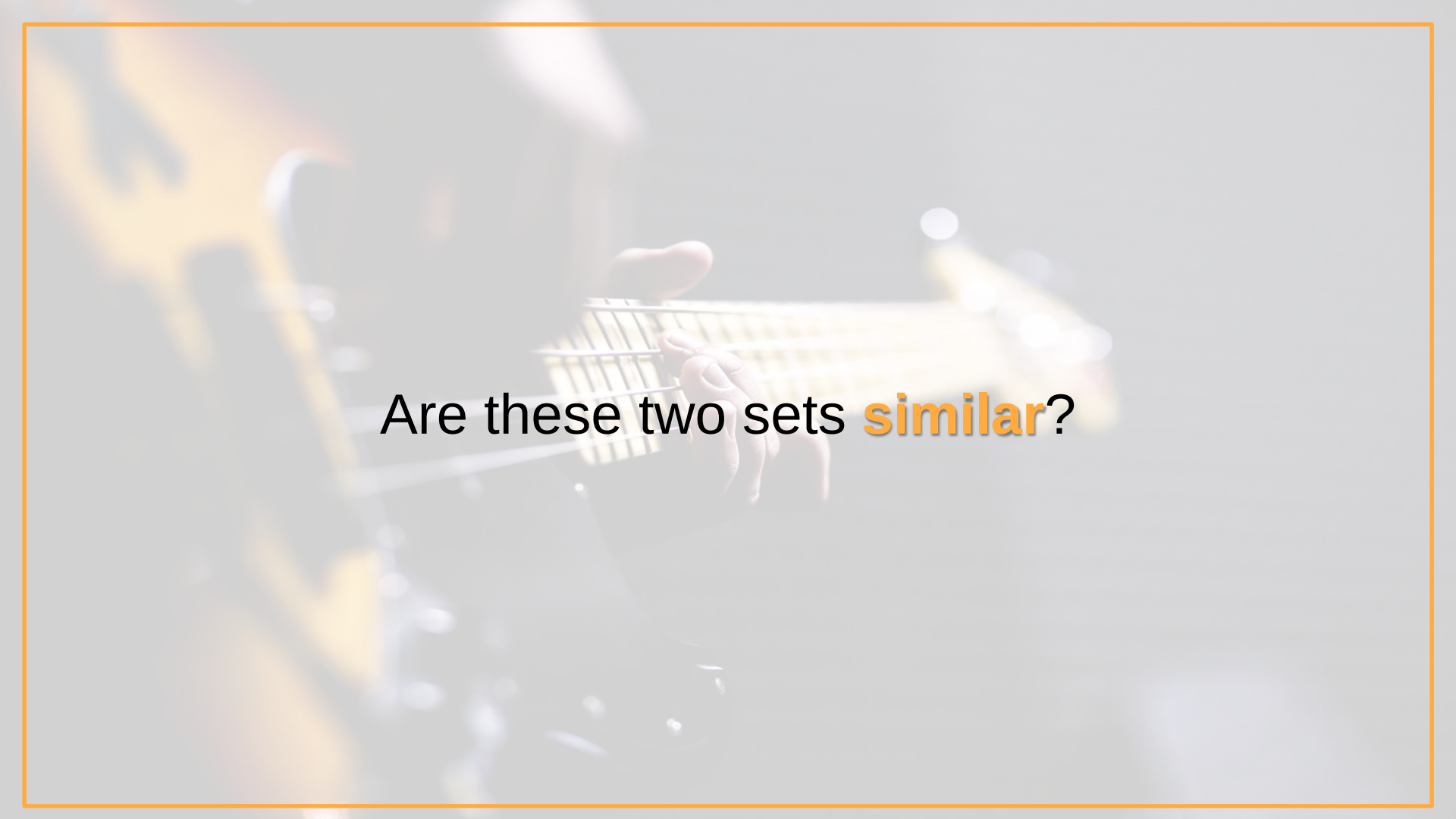
Most frequent tempo: **93.03 – 99.646** bpm
12,650 entries

Stat	Value
Max. Tempo	212.117000
Min. Tempo	46.718000
Mean	121.458016
Median	117.355500



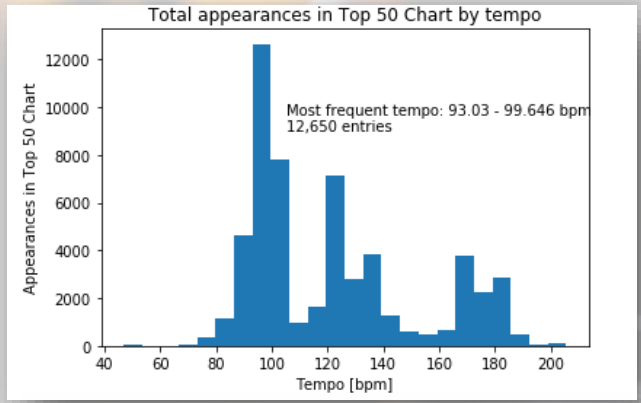
Most frequent tempo: **93.03 – 99.646** bpm
96 entries

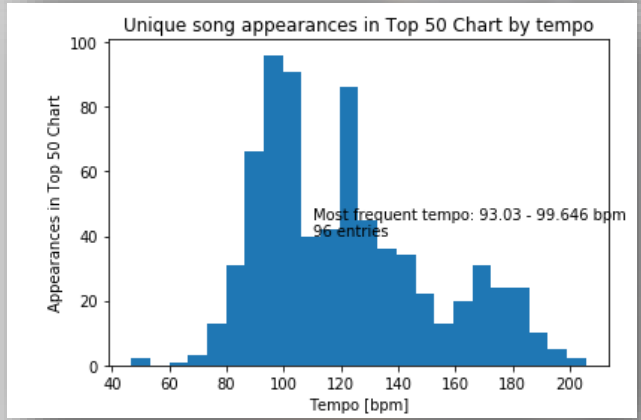
Stat	Value
Max. Tempo	212.117000
Min. Tempo	46.718000
Mean	121.458016
Median	117.355500



Are these two sets **similar**?

Music Tempo Analysis



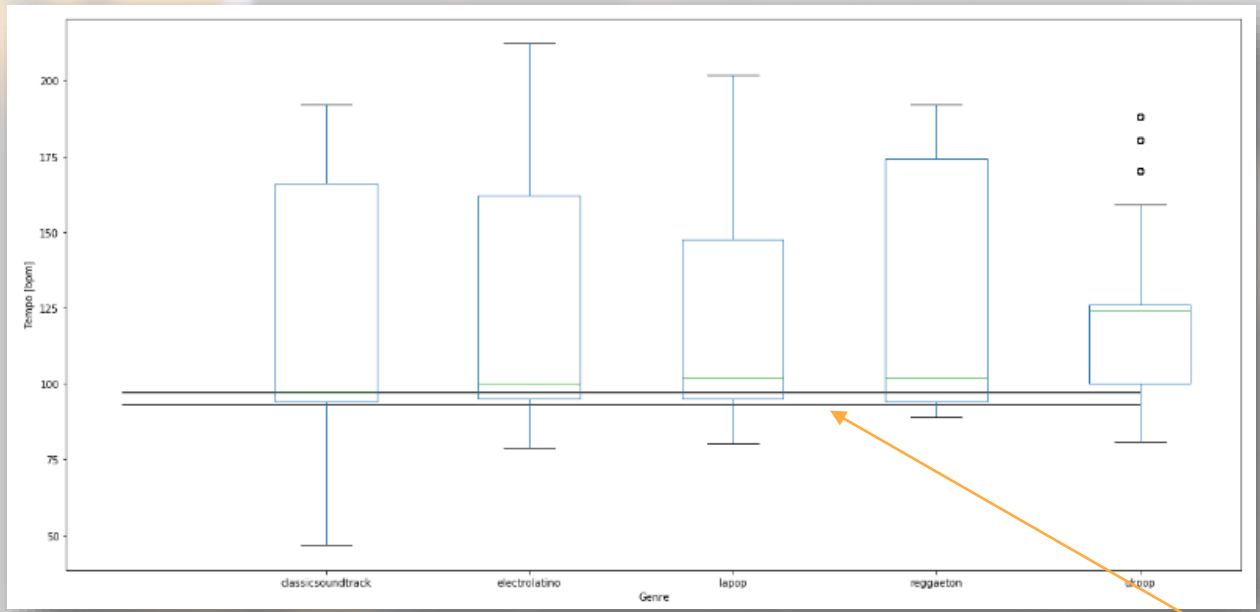


T-test results:

- T-stat: 0.208
- P-value: **0.834**

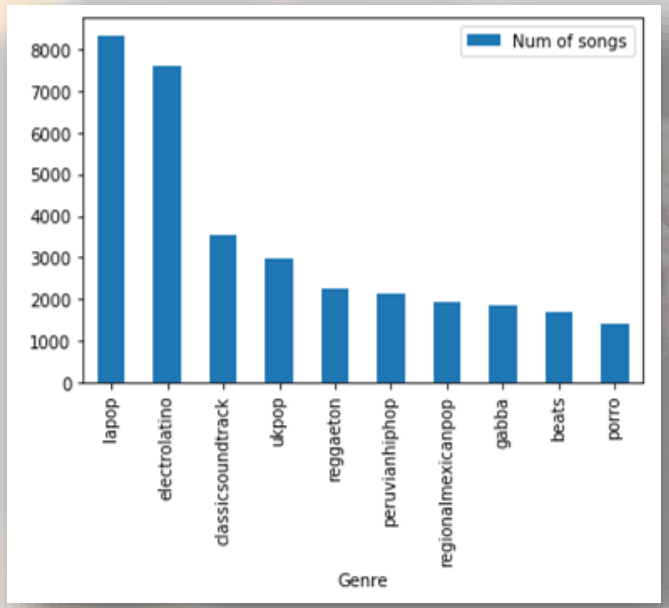
Both sets are similar and it is possible to say that **tempo** is a relevant feature for the popularity of a song!

Music Tempo Grouped by Genre



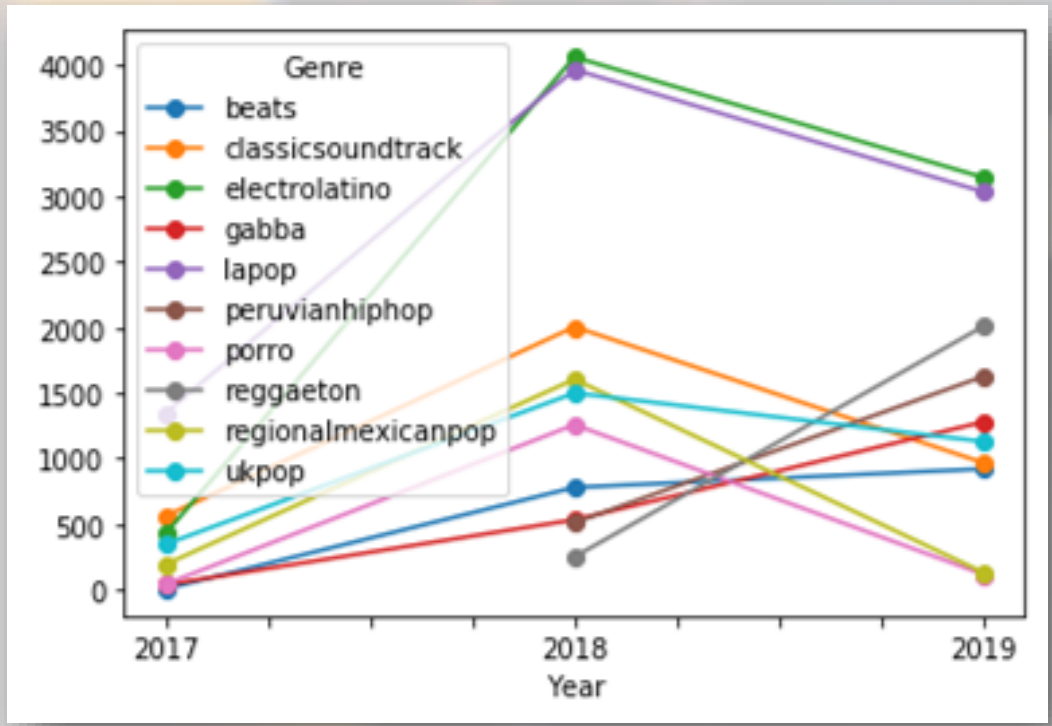
Tempo sweet spot: 93.03 – 99.646 bpm

Top 10 Most Listened Genres



	Genre	Num of songs
0	lapop	8340
1	electrolatino	7626
2	classicsoundtrack	3531
3	ukpop	2974
4	reggaeton	2259
5	peruvianhiphop	2137
6	regionalmexicanpop	1927
7	gabba	1850
8	beats	1707
9	porro	1417

Linear regression of most listened genres



Out of the most popular genres, reggaeton shows the biggest increase in popularity from 2018 to 2019.

Conclusions

- There is no significant correlation between the popularity of a song and its danceability or energy.
- Songs with a tempo between 93 and 100 bmp are more popular. Out of the five most popular genres, classic soundtrack and electrolatino fall closest to this range, followed by lapop and reggaeton .
- Reggaeton genre is more likely to be in top 50 songs amongst the next year due to the high increase between 2018 and 2019.


Thank you!