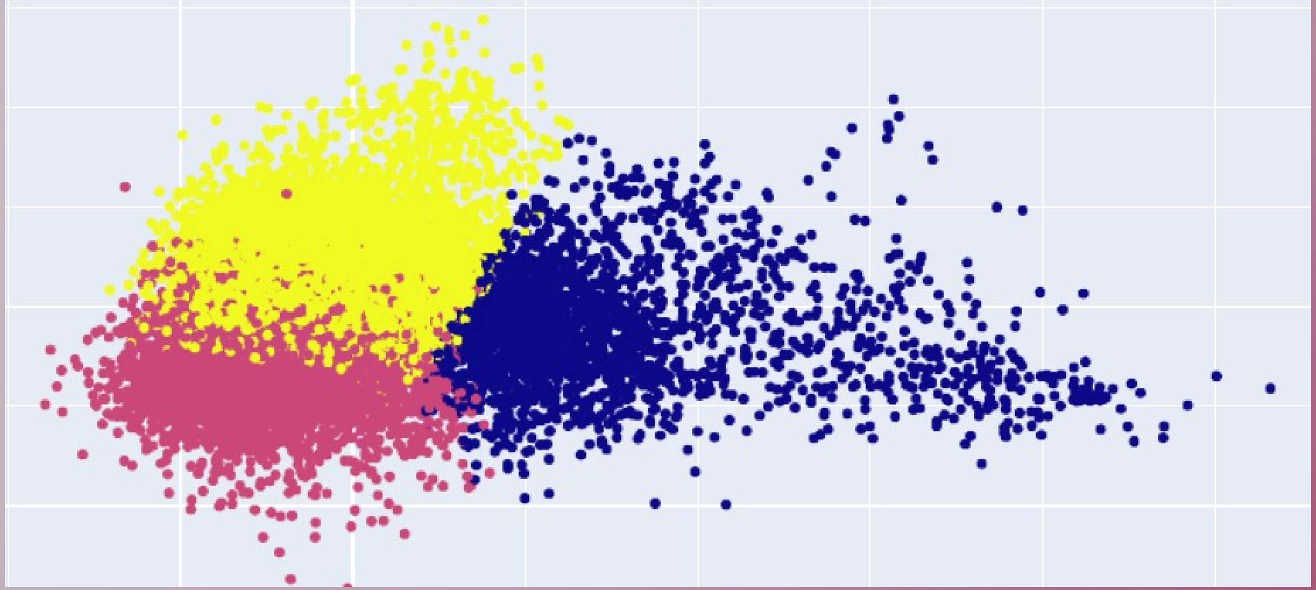
The Recipe for a Popular Song:

Spotify Song Cluster Analysis



February 2021
Team 65
Mark McKenzie, Edna Fernandes, Saba Asefa

Our Goal:

What features make a popular song?

Identify the components of popular songs on Spotify

-What trends exist?

-How have they changed over time?

Make predictions about popularity given a song's audio features

-Create a recipe for popular songs

-Increase artist exposure through song recommendations

Our Approach:

Identify Characteristics of Popular Songs

What are the most important features associated with a songs popularity?

Cluster Characteristics

Create groupings of related songs

-Identify the features that differentiate the clusters

-Isolate the features that make popular cluster popular

Get to Know About the Variables..

Data Details

The dataset contains a nearly 14k song subset of a Kaggle dataset sourced from Spotify's web API. Songs in the dataset were released between 2014 and 2020 .

Spotify songs are rated for their audio features which help with create recommendations of songs a user may like based on their current selection. These audio features are included as variables for the songs in our dataset.

Songs were rated on 10 audio features, assigned a popularity rating, and additionally categorized by year, key, and artist.

Select a variable

Variable: Acousticness (1) ▾

Minimum Value

0

Maximum Value

1

Description

A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.

Descriptive Analysis:

Audio Feature Distributions

-How have trends changed over time?

-How are songs distributed by each of the audio features?

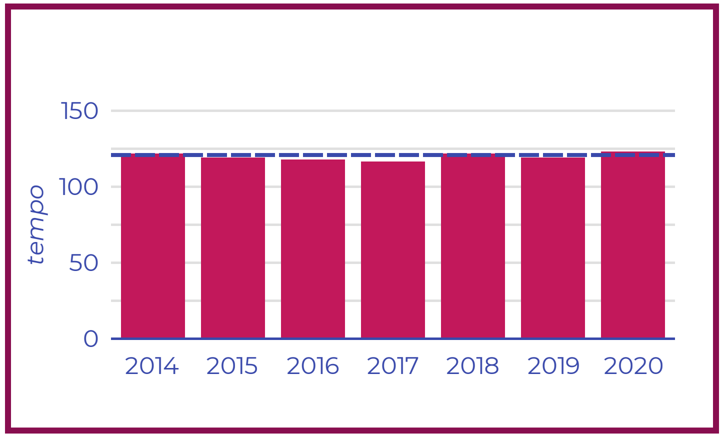
-Are popular songs distributed differently?

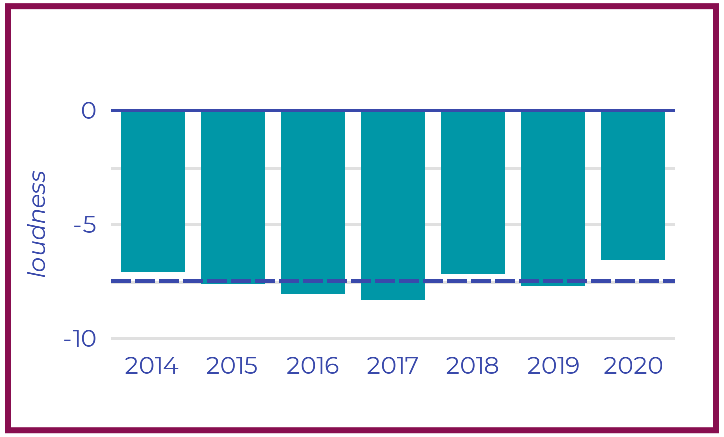
Interactions with Popularity

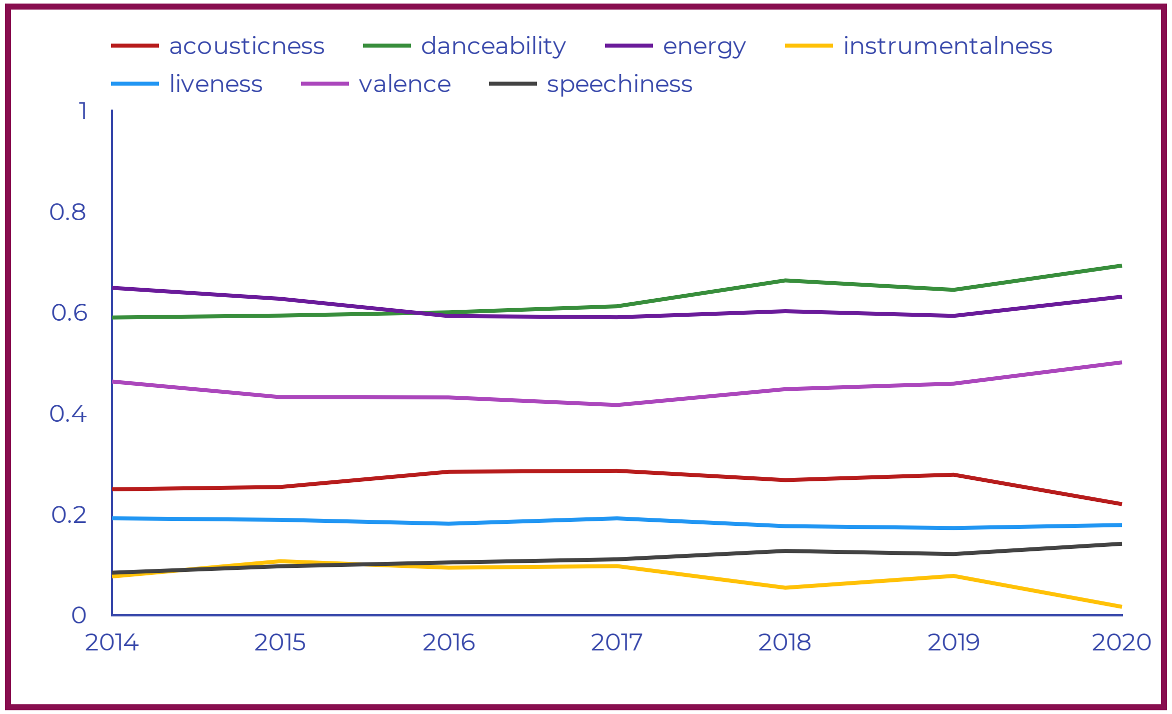
-What features appear to be related to a songs popularity

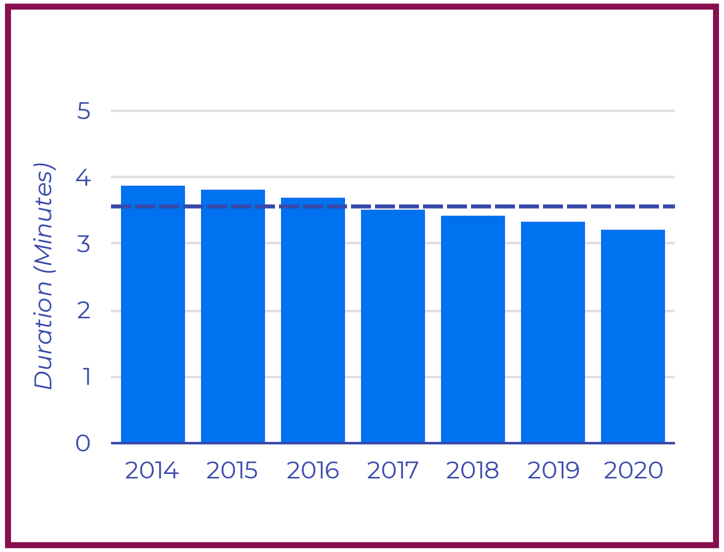
Music Trends Over the Years

Since 2014, songs have generally gotten ***louder***, have shorter ***durations*** and have higher ***tempos***. ***Danceability*** and ***Speechiness*** have also increased while ***Insturmentalness*** has decreased.







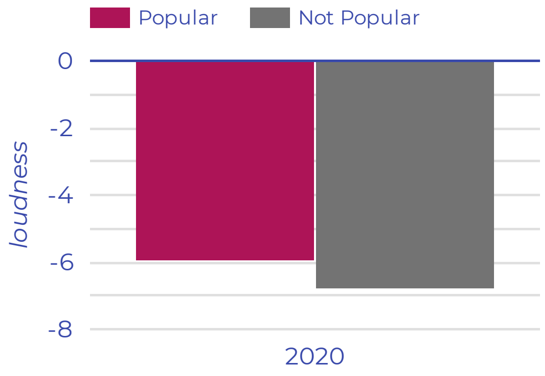


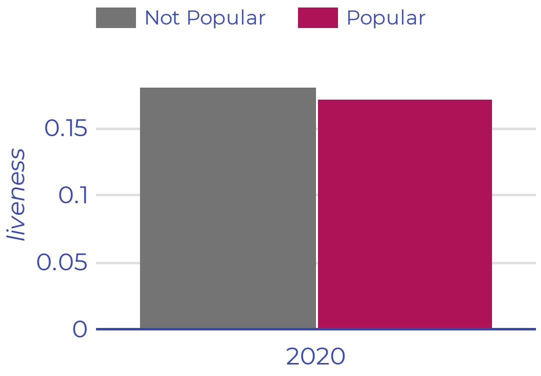
Components of a Popular Song

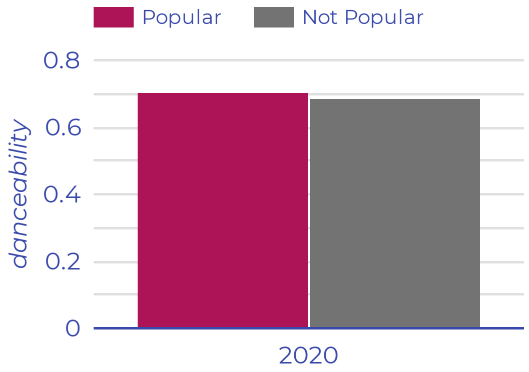
Select a Year

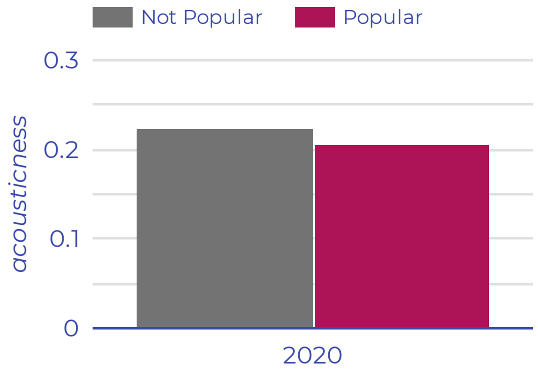
year: 2020 (1) ▾

Songs released that were popular in 2020 were similarly distributed to all other songs in all audio features except ***Loudness***, ***Danceability***, ***Acousticness***, and ***Liveness***. Popular songs were louder less speech-like (speechiness). Popular songs also appeared to be slightly more danceable and slightly less acoustics.









	name	artists	popularity ▾	instrumentalness	speechiness	loudness	valence
1.	Dakiti	Bad Bunny', 'Jhay Cort...	100	0.0	0.1	-10.1	0.1
2.	Mood (feat. iann dior)	24kGoldn', 'iann dior	99	0.0	0.0	-3.6	0.8
3.	WAP (feat. Megan The...	Cardi B', 'Megan Thee ...	96	0.0	0.4	-7.5	0.4
4.	What You Know Bout ...	Pop Smoke	96	0.0	0.4	-8.5	0.5
5.	Blinding Lights	The Weeknd	96	0.0	0.1	-5.9	0.3
6.	Holy (feat. Chance The...	Justin Bieber', 'Chance...	95	0.0	0.4	-8.1	0.4
7.	Lonely (with benny bl...	Justin Bieber', 'benny ...	95	0.0	0.0	-7.1	0.1
8.	you broke me first	Tate McRae	95	0.0	0.1	-9.4	0.1
9.	Lemonade	Internet Money', 'Gunn...	94	0.0	0.1	-6.2	0.5
10.	Relación - Remix	Sech', 'Daddy Yankee', '...	94	0.0	0.1	-3.4	0.8

Modeling & Statistical Analysis:

Cluster Analysis

-How should songs be grouped to uncover common characteristics among popular songs?

-What are the features of these clusters?

ANOVA Testing

-Do the clusters differ significantly by audio feature?

K-Means Cluster Analysis

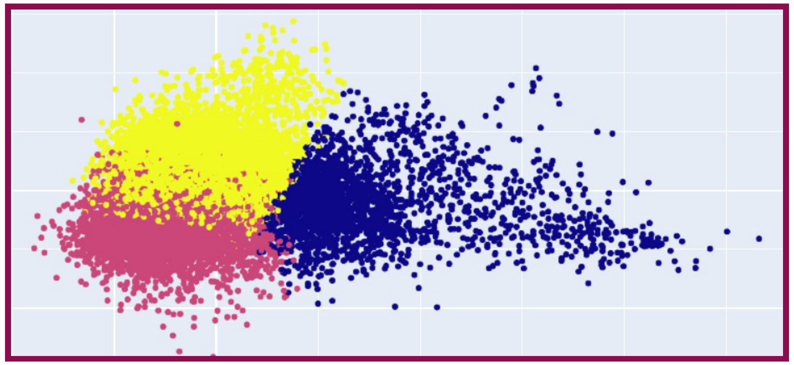
We used K-Means clustering to create three clusters of songs that were more similar to other songs in the same cluster and more distinct from songs in other clusters.

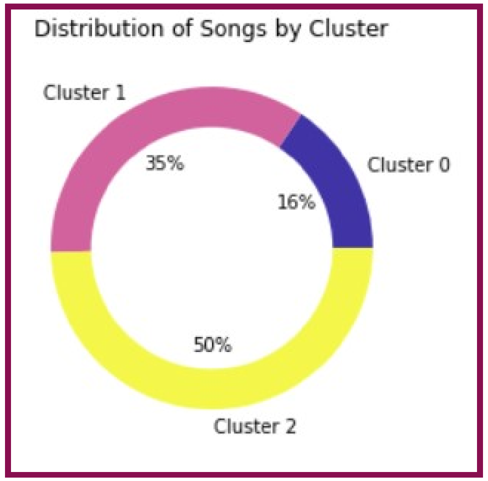
We determined three clusters to be the optimal number using the silhouette method.

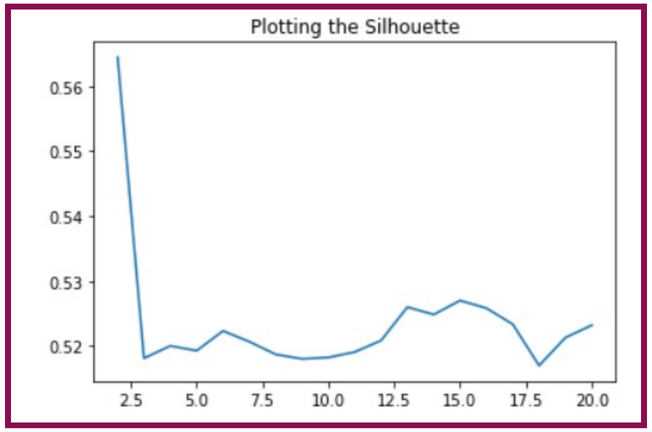
Nearly half of the songs fell into cluster two

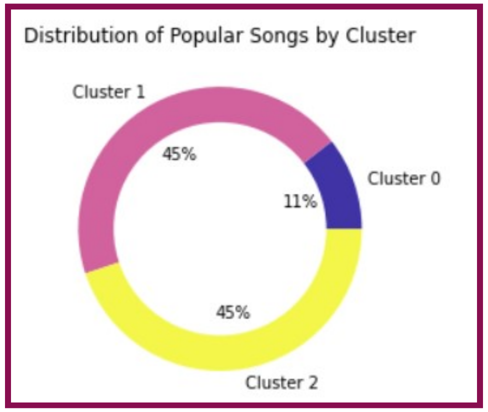
An equal share (45%) of popular songs fell into clusters 1 and 2.

Cluster 1 seemed to have more popular songs than would be expected given the number of songs assigned to the cluster.









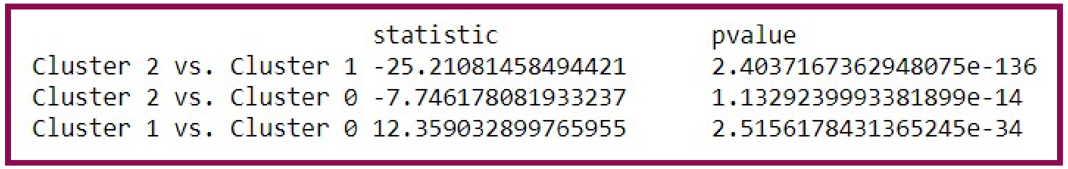
ANOVA Testing

The **Analysis of Variance (ANOVA)** test is used to determine if there is a significant difference between three or more groups along some numeric value. We used this test to determine if popularity differs significantly between the song clusters as it appears to based on distributions.

Null Hypothesis: There is no difference in popularity between the three song clusters.

Alternative Hypothesis: At least one of the clusters has differs in popularity from the others.

P-value: 0.05/6 --> 0.0083



```
                        statistic               pvalue
Cluster 2 vs. Cluster 1 -25.21081458494421      2.4037167362948075e-136
Cluster 2 vs. Cluster 0 -7.746178081933237      1.1329239993381899e-14
Cluster 1 vs. Cluster 0 12.359032899765955      2.5156178431365245e-34
```

✓

We ***rejected*** our null hypothesis and concluded there is a significant difference in mean popularity among the clusters.

ANOVA Testing - Other Variables

We tested the other audio features that were most highly correlated with popularity (***duration, speechiness, loudness***, ***liveness***, and ***valence***) for significant differences between the clusters.

Null Hypothesis: There is no difference in these features between the three clusters.

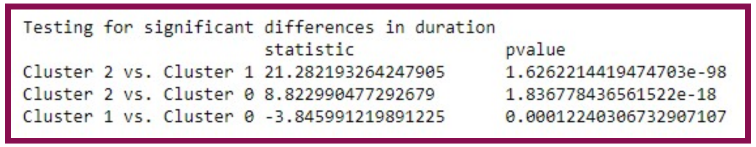
Alternative Hypothesis: At least one of the clusters has differs from the others.

P-value: 0.05/6 --> 0.0083

We ***rejected*** our null hypothesis and concluded there is a significant difference in mean ***duration, loudness speechiness,*** *and* ***valence*** between the three clusters.

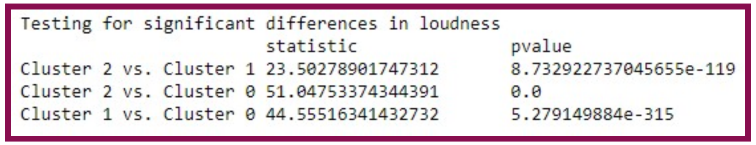
We failed to reject our null hypothesis finding no significant difference in ***liveness*** in our clusters.

Duration ✓



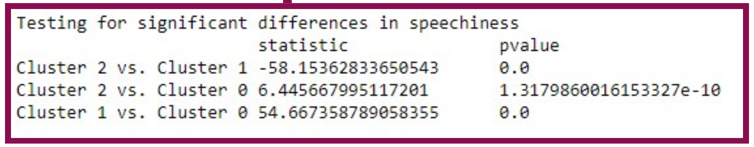
```
Testing for significant differences in duration
                        statistic               pvalue
Cluster 2 vs. Cluster 1 21.282193264247905      1.6262214419474703e-98
Cluster 2 vs. Cluster 0 8.822990477292679       1.836778436561522e-18
Cluster 1 vs. Cluster 0 -3.845991219891225      0.00012240306732907107
```

Loudness ✓



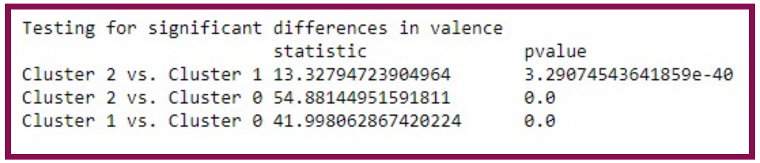
```
Testing for significant differences in loudness
                        statistic               pvalue
Cluster 2 vs. Cluster 1 23.50278901747312       8.732922737045655e-119
Cluster 2 vs. Cluster 0 51.04753374344391       0.0
Cluster 1 vs. Cluster 0 44.55516341432732       5.279149884e-315
```

Speechiness ✓



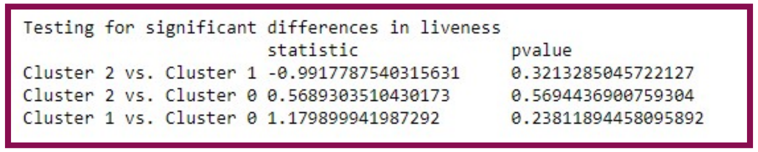
```
Testing for significant differences in speechiness
                        statistic               pvalue
Cluster 2 vs. Cluster 1 -58.15362833650543      0.0
Cluster 2 vs. Cluster 0 6.445667995117201       1.3179860016153327e-10
Cluster 1 vs. Cluster 0 54.667358789058355      0.0
```

Valence ✓



```
Testing for significant differences in valence
                        statistic               pvalue
Cluster 2 vs. Cluster 1 13.32794723904964       3.29074543641859e-40
Cluster 2 vs. Cluster 0 54.88144951591811       0.0
Cluster 1 vs. Cluster 0 41.998062867420224      0.0
```

Liveness 🚫



```
Testing for significant differences in liveness
                        statistic               pvalue
Cluster 2 vs. Cluster 1 -0.9917787540315631     0.321328504572212 7
Cluster 2 vs. Cluster 0 0.5689303510430173      0.5694436900759304
Cluster 1 vs. Cluster 0 1.179899941987292       0.23811894458095892
```

Findings and Recommendations:

We found that the audio features that had the highest impact on a song's popularity were the song's ***acousticness, danceability,*** and ***duration.*** There was a positive relationship between danceability and popularity and a negative one between both acousticness and duration and popularity.

For artists looking to increase exposure, or gain popularity through recommendations on the platform:

Make songs that have similar make-ups to Cluster 1.

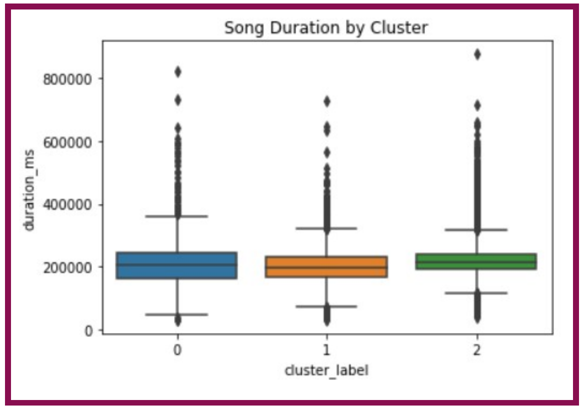
Duration: Make shorter songs (average duration less than 3.5 minutes)

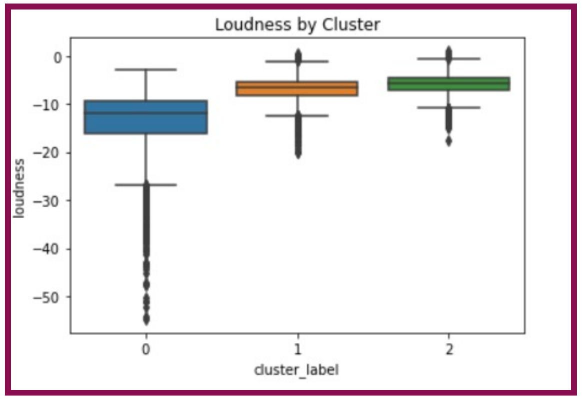
Speechiness: Have a good mix of words and music but emphasize the music (average speechiness 0.2)

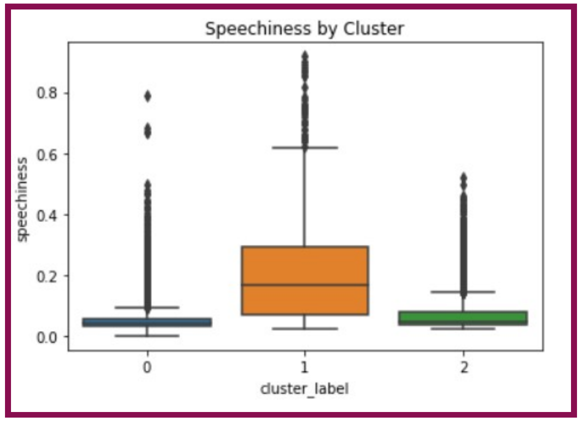
Loudness: Songs should be on the louder side (average loudness -6.8)

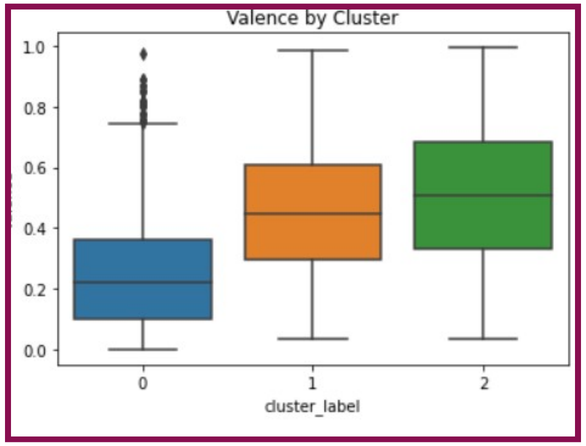
Valence: Songs should be moderately positive (average valence 0.5).

These were the audio features that were most highly and significantly associated with popularity.









Most Popular Artists

2020's Most Popular Artists

	artists	name ▾	popularity
1.	BTS	16	81.88
2.	Ariana Grande	11	84.5
3.	Juice WRLD	11	80
4.	Taylor Swift	7	77.71
5.	BLACKPINK	6	80.83
6.	Pop Smoke	6	83.17
7.	Bad Bunny	6	81.5
8.	The Kid LAROI	5	80.83
9.	Ava Max	5	82.2
10.	The Weeknd	5	83.8

These are Spotify's ten artist with the most popular songs that were released in 2020 along with the average popularity of those songs. Select an ***artist*** to ***drilldown*** into their most popular songs.

Who Are People Still Listening To?

	artists	name ▾	popularity
1.	Billie Eilish	19	80.11
2.	Harry Styles	16	81.11
3.	XXXTENTACION	12	81.83
4.	Post Malone	11	80.45
5.	Juice WRLD	11	81.73
6.	Ariana Grande	9	81.22
7.	Ed Sheeran	8	81.75
8.	Frank Ocean	7	78.43
9.	One Direction	7	78.43
10.	Drake	6	79.17

These are the 10 artists with the most songs that people are still listening to in 2020 (popular in 2020). Select an ***artist*** to ***drilldown*** into their most popular songs.