

Beckett Dumo

Dr. Zietz

12/8

INFO 2201

My question for my project was, do radio stations play the Christmas music people want to listen to? I planned to answer it using the Spotify API for song streaming data, a spreadsheet on the Billboard top 100 from 1958 to 2017, and a radio API on radio data. I wanted to use the streaming data and billboard data to find the most popular streaming songs and then cross-reference them with the radio data to see correlations between streaming services and radio in songs. I first began by importing all the necessary libraries into a Jupyter Notebook, pandas, requests, etc. Next, I accessed the Spotify API with my access token and made preliminary requests to understand the API. I found out early on that Spotify doesn't allow API requests on monthly listener data; it instead generates a popularity score out of 100 based on a song or artist based on their internal listener data. To overcome this, I decided to use the popularity score of each track instead, as the scores listed were similar to what I expected to see with the monthly statistics. I then imported my Billboard data using pandas and made a value count of songs to see the top five most prevalent songs in the dataset. The songs were Jingle Bell Rock, All I want for Christmas is you, Rockin' around the Christmas Tree, White Christmas, and The Chipmunk Song. Using this data, I made Spotify API requests for each song. Next, I created a data frame and filled it with the popularity score of each piece and the music itself.

Spotify Popularity Score is a 0-100 score of how popular a track is compared to other tracks at the current time, 0 being the least popular and 100 being the most

	Song	Spotify Popularity
0	JINGLE BELL ROCK	90
1	ALL I WANT FOR CHRISTMAS IS YOU	94
2	ROCKIN' AROUND THE CHRISTMAS TREE	92
3	WHITE CHRISTMAS	82
4	THE CHIPMUNK SONG	53

Unfortunately, the radio API I planned on using went out of service, so I switched my third data source to a second CSV. I found a top 100 list created by WCBS FM, a New York Classics station, and imported it using pandas. I compared the top five radio songs to the top five Billboard songs and discovered both lists top five shared three songs.

JINGLE BELL ROCK	28
ALL I WANT FOR CHRISTMAS IS YOU	20
ROCKIN' AROUND THE CHRISTMAS TREE	19
WHITE CHRISTMAS	16
THE CHIPMUNK SONG (CHRISTMAS DON'T BE LATE)	16

Name: song, dtype: int64

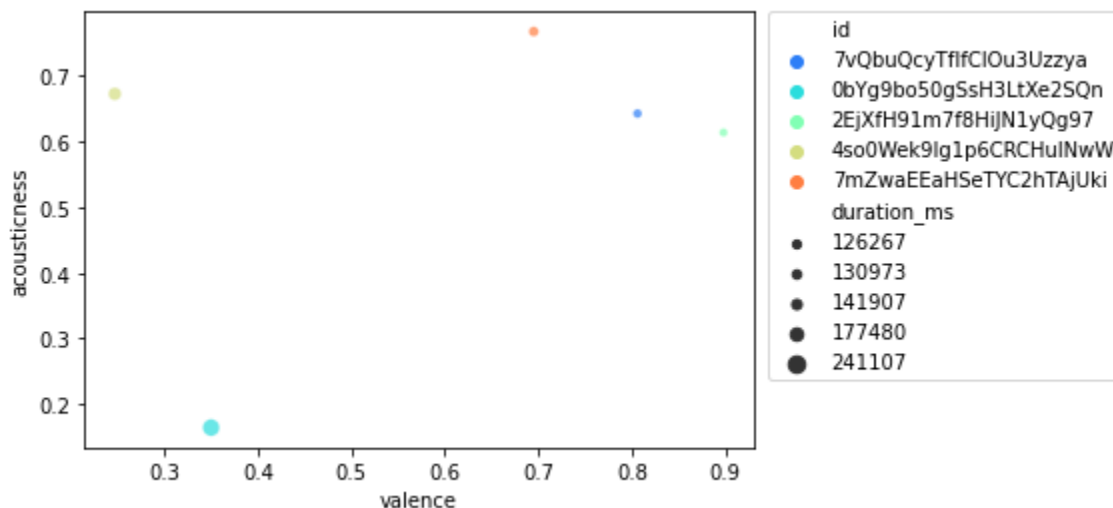
Billboard is above, and radio station is below.

	Rank	Song	Artist
0	1	White Christmas	Bing Crosby
1	2	The Chipmunk Song	The Chipmunks
2	3	Rudolph, The Red Nosed Reindeer	Gene Autry
3	4	I Saw Mommy Kissing Santa Claus	Jimmy Boyd
4	5	Jingle Bell Rock	Bobby Helms

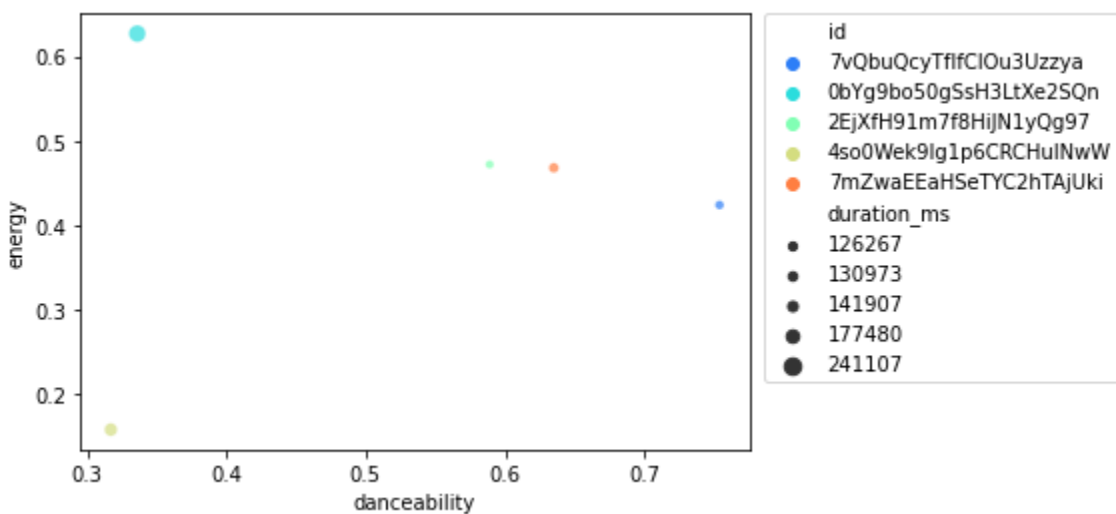
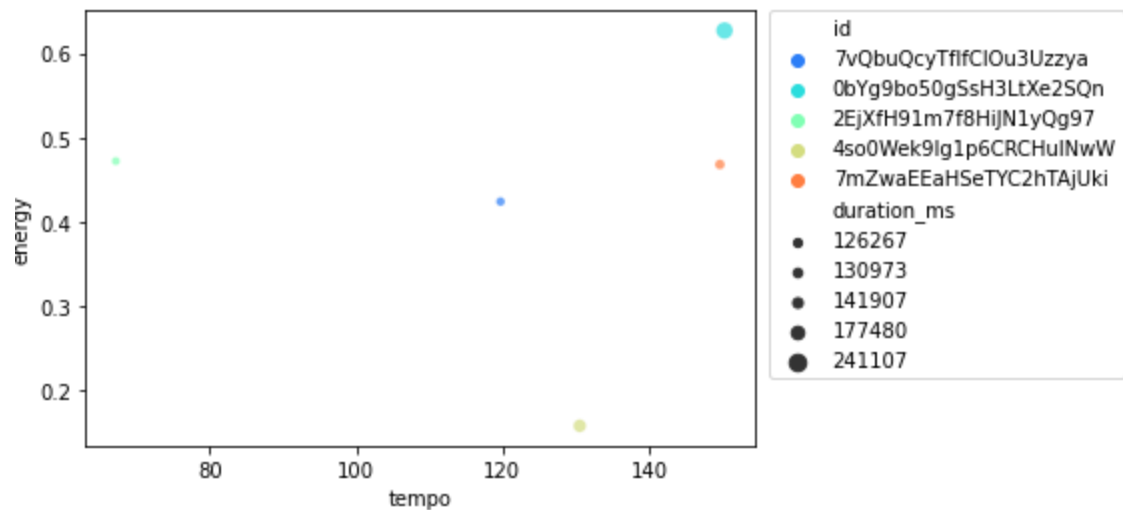
While this discovery meant my main question was answered, it also spurred me into exploring the datasets further. I explored two new questions. First, Spotify generates analytical data on songs within their API, displaying their valence, energy, danceability, etc. This got me thinking about what traits popular Christmas songs share, and is there a “template” that they share? To begin data collection, I made an API request for the audio features of the top 5 songs in the Billboard set and organized them into a pandas data frame. This looked like this:

	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo
0	0.754	0.424	2	-8.463	1	0.0363	0.643	0.00000	0.0652	0.806	119.705
1	0.336	0.627	7	-7.463	1	0.0384	0.164	0.00000	0.0708	0.350	150.273
2	0.589	0.472	8	-8.749	1	0.0502	0.614	0.00000	0.5050	0.898	67.196
3	0.317	0.158	9	-16.485	1	0.0381	0.673	0.35300	0.3380	0.247	130.503
4	0.635	0.468	3	-7.175	1	0.0417	0.768	0.00134	0.5760	0.695	149.640

I began using seaborn to make scatter plots to examine how each variable correlates with the other. The first I looked at was acousticness, a scale of 0-1 on how much of the sound originates from acoustic instruments and how it correlates with valence, a scale of 0-1 on how happy the song is.

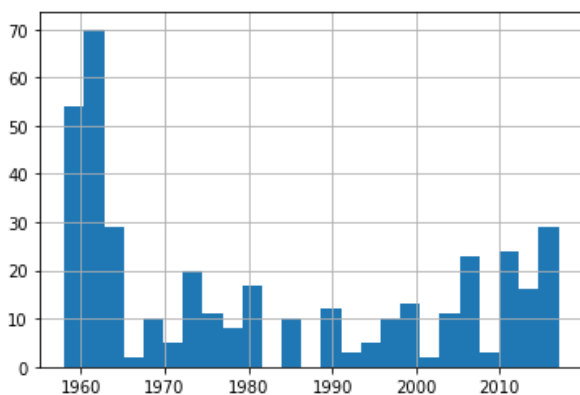


The colors from top to bottom are the songs Jingle Bell Rock, All I want for Christmas is you, Rockin’ around the Christmas Tree, White Christmas, and The Chipmunk Song, respectfully. As you can see, All I want for Christmas is you is an example of low valence and acoustics. But, on the other hand, Rockin’ around the Christmas Tree is an example of high valence and acoustics.

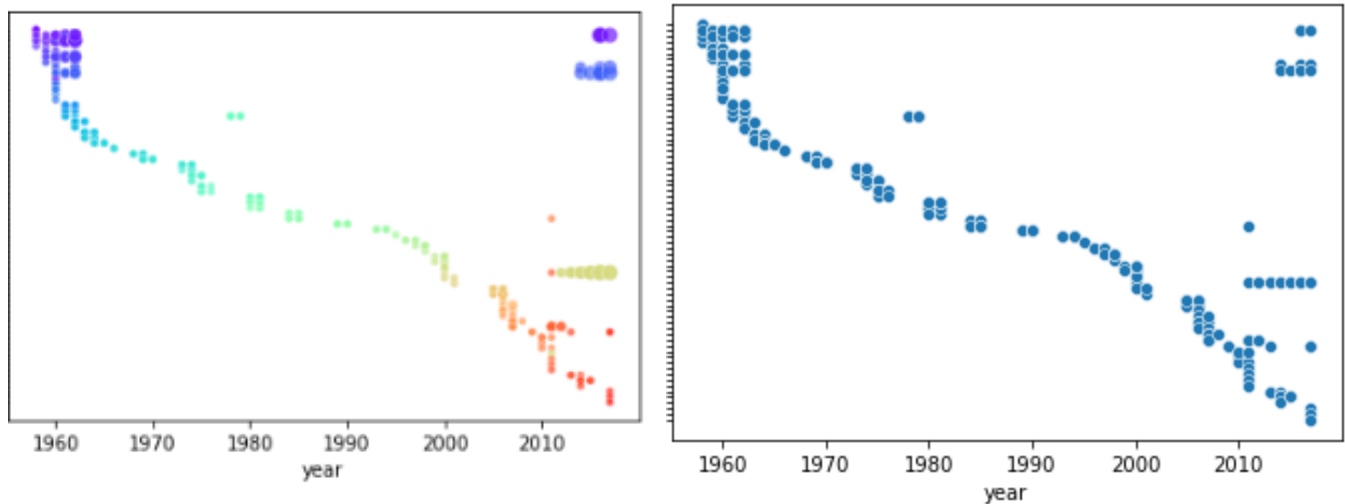


These two scatterplots looked at how energy correlated with a song's tempo and danceability. The plots show that there doesn't seem to be a perfect template for a popular Christmas song. I equated this to how White Christmas and Jingle Bell Rock, two popular carols, are very different in tempo, valence, energy, etc.

The second question I explored was what Era made the most and best Christmas songs. Using the Billboard data, I imported earlier, I created this histogram of songs by year.



This histogram shows an exciting discovery, the large majority of Christmas songs in the top 100 billboard dataset were from the 60s. I made a scatterplot in response to the histogram and added the variable instance to see if songs from the 60s were billboard top 100 of different decades.



Both charts make a scatterplot where x is the year and y is the song. The plot on the left additionally has a hue set to the song's artist, size as the instance to show how many times the same song appeared in the data, and a rainbow palette to differentiate the pieces. The plots show that several songs from the 60s also appear on the billboard top 100s list in the late 2010s. My data concluded that radio stations and Spotify listeners agree on many of the same songs. Most popular Christmas songs were created in the 60s and are still listened to today. However, popular Christmas songs don't share similar features.

If I continue working on this topic, I will use a third-party API to access music streaming monthly listener data for all sites and use it to create visualizations of listener tendencies with Christmas songs and percentage changes in listeners.