THE IMPACT OF COVID-19 ON STREAMING MUSIC LISTENERS' BEHAVIORS: 2019-PRESENT

Project Demo

Jou Lee

BACKGROUND

- The outbreak of COVID-19 and the subsequent global pandemic has radically impacted the life in the U.S and around the world. Music consumption decreased by 12.5 percent after WHO's pandemic declaration on Mar 11, 2020.
- Streaming consumption has fallen compared to 2019 with only a 17% increase at the end of 2020. The industry is now experiencing a change in people's behaviors. The pandemic might reduce music consumption in streaming music by restricting travel and cutting down the music-friendly activities like lockdowns and work from home activities.
- Although increased at-home activities caused an overall loss in streaming music, they also inspired new playlists for cooking, cleaning, and in-door exercising. The music behavior of people during the pandemic and lockdowns reflects the resilience of people, artists, and record companies at this difficult time.

RESEARCH QUESTIONS

What are the changes on streaming music?

By analyzing the changes, we could see the impact of the pandemic on peoples' music preferences. According to some annual reports, children music and country music unexpectedly showed growth during the pandemic.

What is the impact of COVID-19 on people's music preferences?

Would there be significant patterns or differences during the peaks of COVID cases? By measuring the relationship of music trends and confirmed cases, we could know whether a correlation exists between them.

	The LFM-1b Dataset	The Million Song Dataset
Overview	This is a dataset of more than one billion music listening events created by more than 120,000 users of a single platform, Last.fm. It's user level data includes basic demographics and a wide range of additional user descriptors that reflect their music taste and consumption behavior.	It is a freely-available collection of audio features and metadata for a million con-temporary popular music tracks. Attractive features of the Million Song Database include the range of existing resources to which it is linked, and the fact that it is the largest current research dataset in the field.
Data Sources	Last.FM	The Echo Nest Last.fm, MusicBrainz, Second Hand Song, musiXmatch
Time	Jan 2013- Aug 2014	1922-2011
Tracks	1 Billion	1 Million

DATA SOURCES

MUSIC

Billboard Hot100

The Weekly Top100 Charts showed the most popular songs across all genres, ranked by radio airplay audience impressions, as measured by Nielsen Music, sales data as compiled by Nielsen Music, and streaming activity data provided by online music sources.

Spotify API

Spotify provides access to extract music data, including artist, album and track data. The extracted data would become the metadata of the tracks in Billboard Hot 100, including genre and audio features.

COVID19

Our World in Data COVID-19 dataset

This dataset provides the daily cases, deaths, vaccinations, tests, etc. data of the world. The project would focus on the smoothed USA COVID data of daily cases, deaths, vaccinations and tests to compare the correlation of music and the pandemic.

Billboard Hot100

Get Charts

Beautiful Soup web scraping 148 pages of Hot100 charts

- `link, date, rank, artist_name, track_name`

Spotify API

Get Track Metadata

Spotipy access API to extract metadata for the 14800 billboard tracks

- `artist_info: artist_name, artist_genres`
- `track_audio_features: danceability, energy, key, loudness, mode, speechiness, acousticness, instrumentalness, liveness, valence, tempo, type, duration_ms, time_signature,…`

Our World in Data COVID-19 dataset

For Data Analysis

Download data on website. Clean data and keep only USA data.

- `new_cases, new_deaths, fully_vaccinated_people, new_vaccinations, new_tests`

DATASET DEMO

Dataset Overview

Overview of the datasets and their metadata.

- Streaming Music: Billboard Hot100 & Spotify Audio Features
- COVID-19: Our world in Data

Potential Uses Demo

A quick demo on their potential usages and response to the previous research questions.

- Data Visualization
- Data Analysis

DATASET OVERVIEW Streaming Music: Billboard Hot100 & Spotify Audio Features

date	rank	track	artists	main_artist	artist_main_genre	artist_genres	danceability	energy	loudness	speechiness	acousticness	instrumental	liveness	valence	tempo	duration_ms	time_signatu
1/8/19	1	Without Me	Halsey	Halsey	pop	['dance pop',	0.752	0.488	-7.05	0.0705	0.297	9.11E-06	0.0936	0.533	136.041	201661	4
1/8/19	2	Thank U, Nex	Ariana Granc	Ariana Granc	pop	['dance pop',	0.717	0.653	-5.634	0.0658	0.229	0	0.101	0.412	106.966	207320	4
1/8/19	3	Sunflower (S	Post Malone	Post Malone	rap	[]	0.76	0.479	-5.574	0.0466	0.556	0	0.0703	0.913	89.911	158040	4
1/8/19	4	Sicko Mode	Travis Scott	Travis Scott	rap	['rap', 'slap h	0.834	0.73	-3.714	0.222	0.00513	0	0.124	0.446	155.008	312820	4
1/8/19	5	High Hopes	Panic! At The	Panic! At The	rock	['modern roc	0.579	0.904	-2.729	0.0618	0.193	0	0.064	0.681	82.014	190947	4
1/8/19	6	Happier	Marshmello	Marshmello	edm	['brostep', 'e	0.687	0.792	-2.749	0.0452	0.191	0	0.167	0.671	100.015	214290	4
1/8/19	7	Girls Like You	Maroon 5 Fe	Maroon 5	pop	['pop', 'pop r	0.851	0.541	-6.825	0.0505	0.568	0	0.13	0.448	124.959	235545	4
1/8/19	8	Drip Too Har	Lil Baby & Gu	Lil Baby	hip hop	['atl hip hop'	0.897	0.662	-6.903	0.292	0.0852	0	0.534	0.389	112.511	145543	4
1/8/19	9	ZEZE	Kodak Black	Kodak Black	rap	['florida rap',	0.861	0.603	-5.788	0.176	0.0521	0	0.0924	0.504	98.043	228760	4
1/8/19	10	Better Now	Post Malone	Post Malone	rap	['dfw rap', 'n	0.68	0.578	-5.804	0.04	0.331	0	0.135	0.341	145.038	231267	4
1/8/19	11	Eastside	benny blanco	benny blanco	pop	['electropop'	0.56	0.68	-7.648	0.321	0.555	0	0.116	0.319	89.391	173800	4
1/8/19	12	Wake Up In T	Gucci Mane	Gucci Mane	hip hop	['atl hip hop'	0.8	0.578	-5.144	0.0485	0.00381	0	0.359	0.367	143.01	204665	4
1/8/19	13	Wow.	Post Malone	Post Malone	rap	['dfw rap', 'n	0.829	0.539	-7.359	0.208	0.136	1.78E-06	0.103	0.388	99.96	149547	4
1/8/19	14	Youngblood	5 Seconds Of	5 Seconds Of	rock	['boy band', '	0.596	0.854	-5.114	0.463	0.0169	0	0.124	0.152	120.274	203418	4
1/8/19	15	Mo Bamba	Sheck Wes	Sheck Wes	rap	['rap', 'trap',	0.729	0.625	-5.266	0.0315	0.194	0.00986	0.248	0.261	146.034	183907	4
1/8/19	16	Breathin	Ariana Granc	Ariana Granc	pop	['dance pop',	0.568	0.656	-5.413	0.0433	0.0211	1.37E-05	0.213	0.364	100.049	198160	4
1/8/19	17	Money	Cardi B	Cardi B	pop	['dance pop',	0.95	0.59	-6.508	0.29	0.00534	0	0.11	0.219	130.003	183527	4
1/8/19	18	Lucid Dream:	Juice WRLD	Juice WRLD	rap	['chicago rap	0.511	0.566	-7.23	0.2	0.349	0	0.34	0.218	83.903	239836	4
1/8/19	19	Taki Taki	DJ Snake Fea	DJ Snake	pop	['dance pop',	0.842	0.801	-4.167	0.228	0.157	4.82E-06	0.0642	0.617	95.881	212500	4
1/8/19	20	MIA	Bad Bunny F	Bad Bunny	latin rap	['latin', 'regg	0.817	0.539	-6.349	0.0621	0.0141	0.000496	0.099	0.158	97.062	210368	4
1/8/19	21	Going Bad	Meek Mill Fe	Meek Mill	hip hop	['hip hop', 'p	0.889	0.496	-6.365	0.0905	0.259	0	0.252	0.544	86.003	180522	4
1/8/19	22	Leave Me Alc	Flipp Dinero	Flipp Dinero	rap	['melodic rap	0.792	0.743	-2.806	0.0851	0.107	0	0.183	0.742	150.024	195637	4
1/8/19	23	Trip	Ella Mai	Ella Mai	pop	['dance pop',	0.477	0.61	-5.628	0.144	0.225	0	0.107	0.358	79.882	213993	4
1/8/19	24	Speechless	Dan + Shay	Dan + Shay	country	['contempora	0.616	0.438	-5.968	0.0298	0.356	0	0.24	0.386	135.929	213387	4
1/8/19	25	Natural	Imagine Drag	Imagine Drag	rock	['modern roc	0.704	0.611	-6.112	0.0409	0.217	0	0.0812	0.22	100	189467	4
1/8/19	26	I Like It	Cardi B, Bad	Cardi B	pop	['dance pop',	0.816	0.726	-3.998	0.129	0.099	0	0.372	0.65	136.048	253390	4
1/8/19	27	Love Lies	Khalid & Nor	Khalid	pop	['pop', 'pop r	0.708	0.648	-5.626	0.0449	0.0956	0	0.134	0.338	143.955	201707	4
1/8/19	28	Beautiful	Bazzi Featur	Bazzi	pop	['dance pop',	0.638	0.717	-4.722	0.0337	0.346	0	0.105	0.249	100.027	180000	4
1/8/19	29	Shallow	Lady Gaga &	Lady Gaga	pop	['art pop', 'da	0.572	0.385	-6.362	0.0308	0.371	0	0.231	0.323	95.799	215733	4
1/8/19	30	Better	Khalid	Khalid	pop	['pop', 'pop r	0.596	0.552	-10.278	0.097	0.0765	0.334	0.104	0.112	97.949	229320	4
1/8/19	31	She Got The	Luke Combs	Luke Combs	country	['contempora	0.533	0.907	-3.793	0.0406	0.0292	0	0.386	0.7	150.99	183160	4
1/8/19	32	Baby Shark	Pinkfong	Pinkfong	cartoon	['cartoon', "c	0.829	0.886	-1.746	0.112	0.259	0	0.0559	0.777	115.056	80927	4
1/8/19	33	Close To Me	Ellie Gouldin	Ellie Gouldin	pop	['dance pop',	0.575	0.758	-5.029	0.0618	0.0954	0	0.394	0.493	144.107	182623	4
1/8/19	34	You Say	Lauren Daigl	Lauren Daigl	christian	['ccm', 'chris	0.494	0.632	-6.89	0.0342	0.682	0	0.0869	0.0797	147.873	274693	4
1/8/19	35	Uproar	Lil Wayne	Lil Wayne	hip hop	['hip hop', 'n	0.743	0.87	-2.188	0.212	0.0595	0	0.299	0.884	99.079	194184	4
1/8/19	36	A Lot	21 Savage	21 Savage	hip hop	['atl hip hop'	0.837	0.636	-7.643	0.086	0.0395	0.00125	0.342	0.274	145.972	288624	4
1/8/19	37	Best Shot	Jimmie Allen	Jimmie Allen	country	['black amer	0.629	0.464	-8.72	0.0334	0.75	0	0.0853	0.523	156.051	195053	4
1/8/19	38	Swervin	A Boogie Wi	A Boogie Wi	rap	['melodic rap	0.581	0.662	-5.239	0.303	0.0153	0	0.111	0.434	93.023	189487	4

DATASET OVERVIEW Streaming Music: Billboard Hot100 & Spotify Audio Features

Distinct Count of Artist, Track and Genre

	2019	2020	2021
Distinct count of main_artist	261.0	266.0	249.0
Distinct count of Track	613.0	795.0	628.0
Distinct count of artist_main_genre	12.0	9.0	8.0

Audio Features	Description	Range
Danceability	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. A value of 0.0 is least danceable and 1.0 is most danceable.	0~1
Energy	Energy represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.	0~1
Valence	It describe the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).	0-1

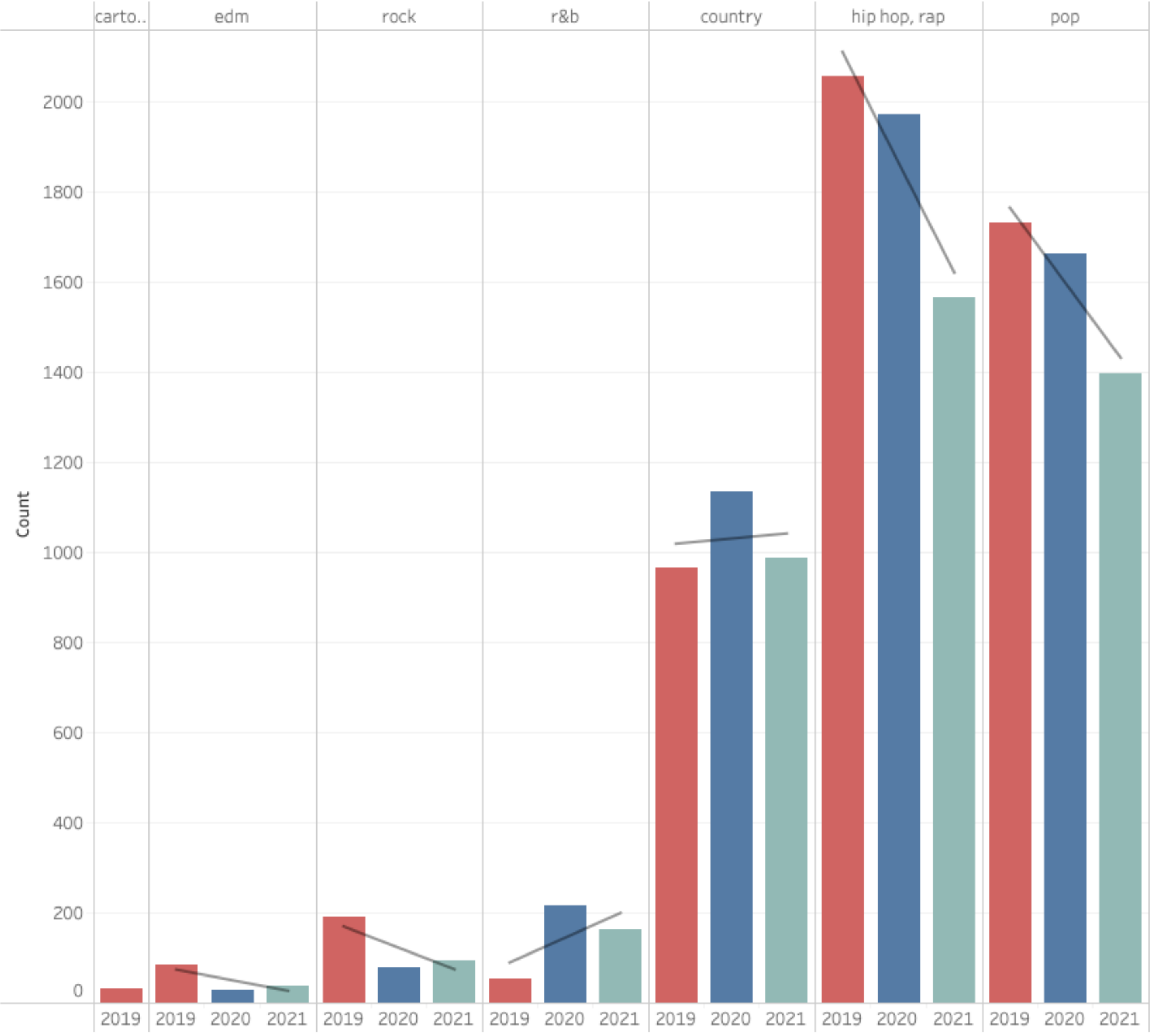
DATASET OVERVIEW COVID-19: Our world in Data

date	new_cases_smoothed	new_deaths_smoothed	new_tests_smoothed	people_fully_vaccinated	new_vaccinations_smoot
11/27/21	585846	6457.143	2371505	583992281	3172759
11/20/21	628977.001	8077.144	9787496	1359674815	9435497
11/13/21	537589.43	8254.713	9964230	1353074711	9187725
11/6/21	513861.144	8487.286	9459732	1345890060	8755426
10/30/21	506651.428	9922.572	9625143	1337873114	8073951
10/23/21	537157	11953.715	10093237	1329425431	5217729
10/16/21	606876.571	11470.142	10228735	1320635017	5377994
10/9/21	693726.572	12378.429	10841007	1310638029	6089756
10/2/21	787512.857	14172.429	11107617	1298944730	6028154
9/25/21	896420.287	14505	11551791	1285893041	4492885
9/18/21	1073305.858	13728	11841326	1270447736	4922294
9/11/21	1030218.572	11165.285	10682164	1253354017	4699892
9/4/21	1146752.142	10451.714	11571593	1237667152	5821449
8/28/21	1079995.572	8511.286	10570331	1219199939	5907858
8/21/21	1008615.572	6317.428	9561432	1201301572	5530208
8/14/21	889860.144	4318.429	8632268	1185850544	4943260
8/7/21	699755.285	3216	7294365	1173413698	4652714
7/31/21	488266.715	2304.143	6007617	1162627395	4198705
7/24/21	312243.428	1842.428	5022811	1152355933	3659164
7/17/21	197151.285	1826.856	4349219	1141432842	3357951
7/10/21	115520.713	1593.285	3701483	1129562849	3185606
7/3/21	89700.286	1811.142	4056509	1117551085	3860817
6/26/21	83489.858	2153	4330741	1102411045	4560428
6/19/21	87181.144	2202	4607316	1083152279	6145717
6/12/21	103153.715	2883.57	5099171	1054131965	7369865
6/5/21	109694.715	3529.999	4931832	1019122281	6521043
5/29/21	159018.285	4286.001	5848241	991514418	9629378
5/22/21	201505.285	4011.428	6596361	953641190	12894575
5/15/21	251209.715	4326.429	7298736	902490249	12918457
5/8/21	317422.858	4909.001	7901589	841336116	15165318

DEMO What are the changes on streaming music?

Data Visualization Tableau

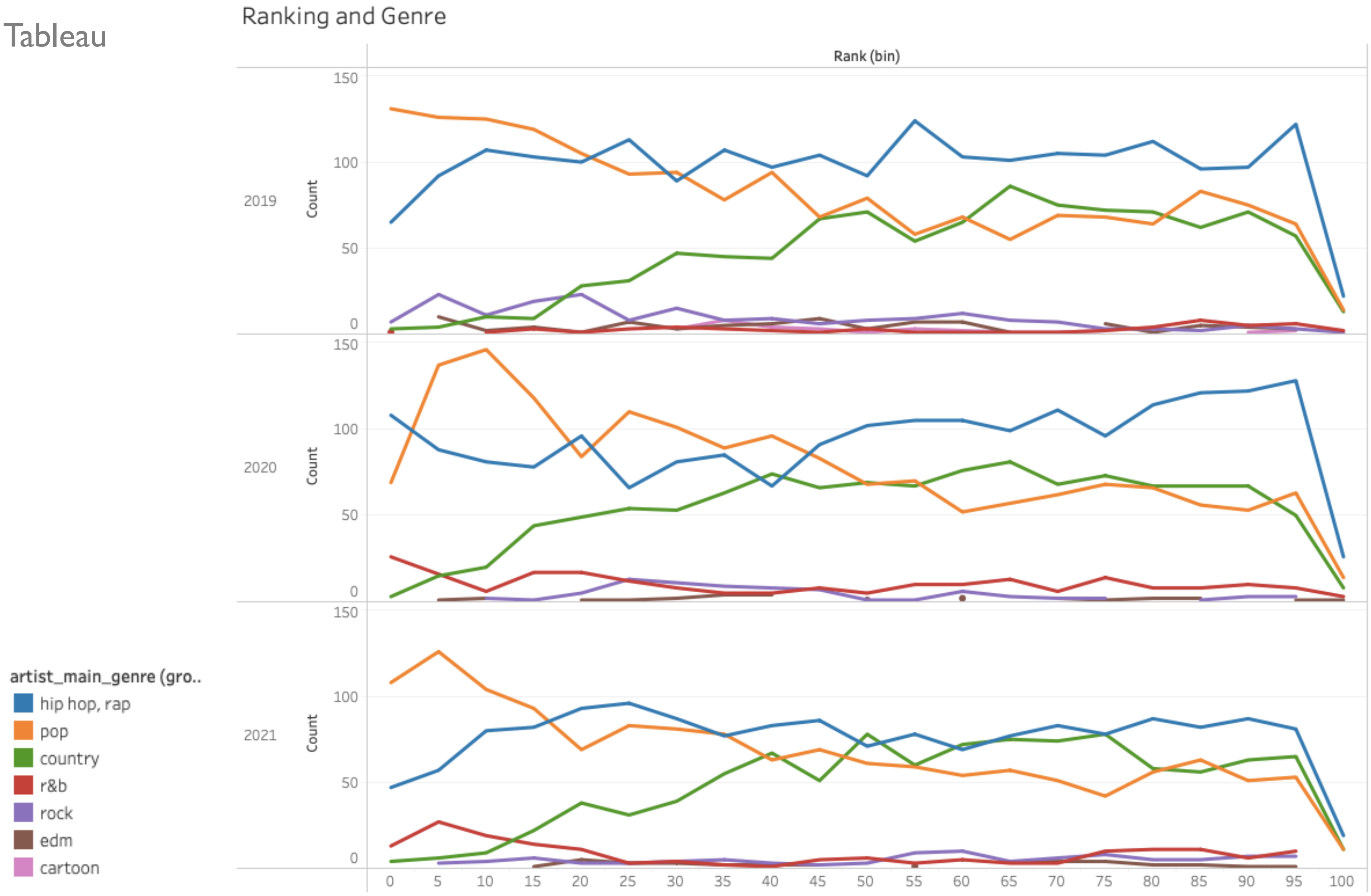
- **Main Genres**
- Rankings
- Audio Features



DEMO What are the changes on streaming music?

Data Visualization Tableau

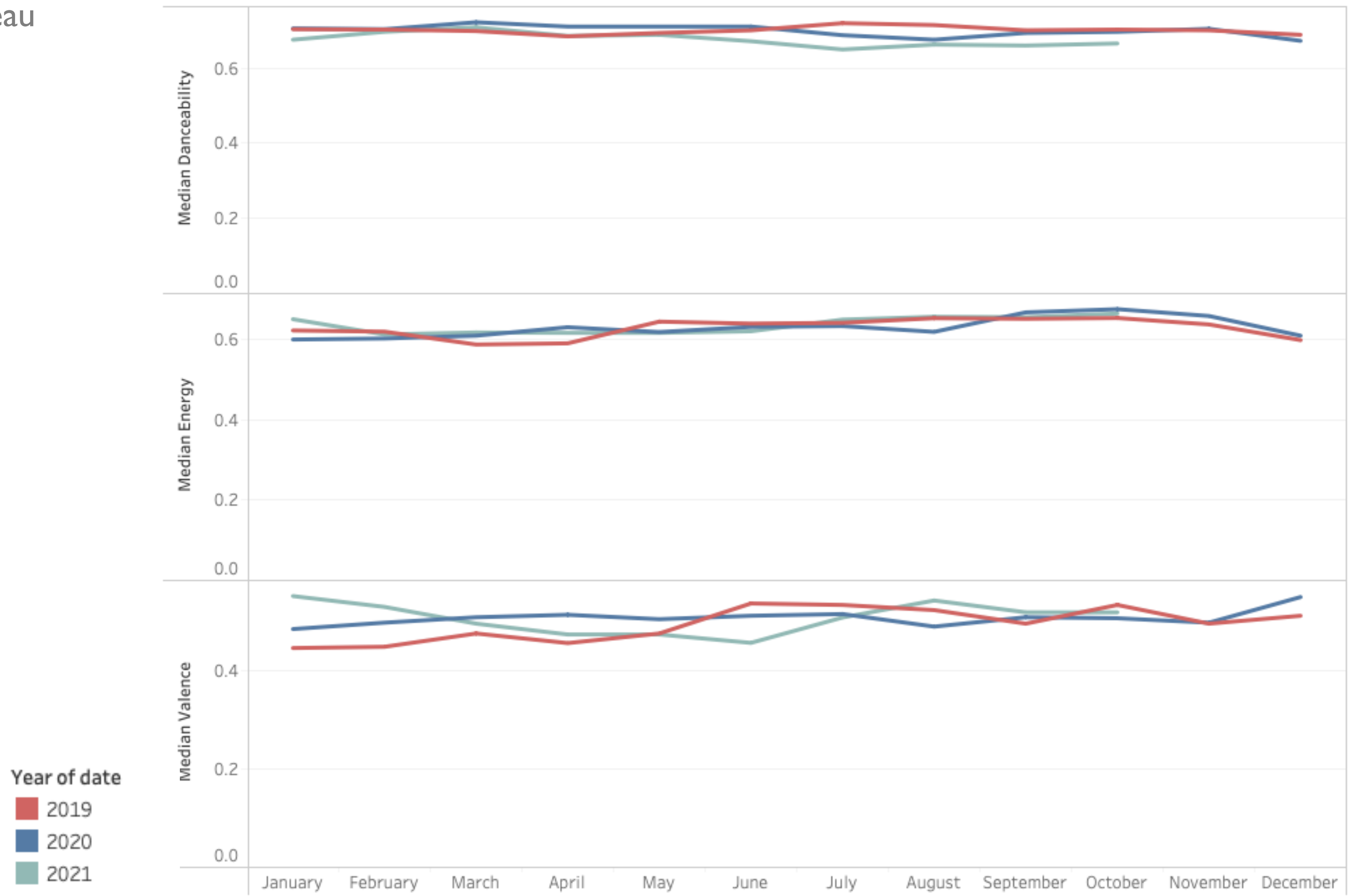
- Main Genres
- **Rankings**
- Audio Features



DEMO What are the changes on streaming music?

Data Visualization Tableau

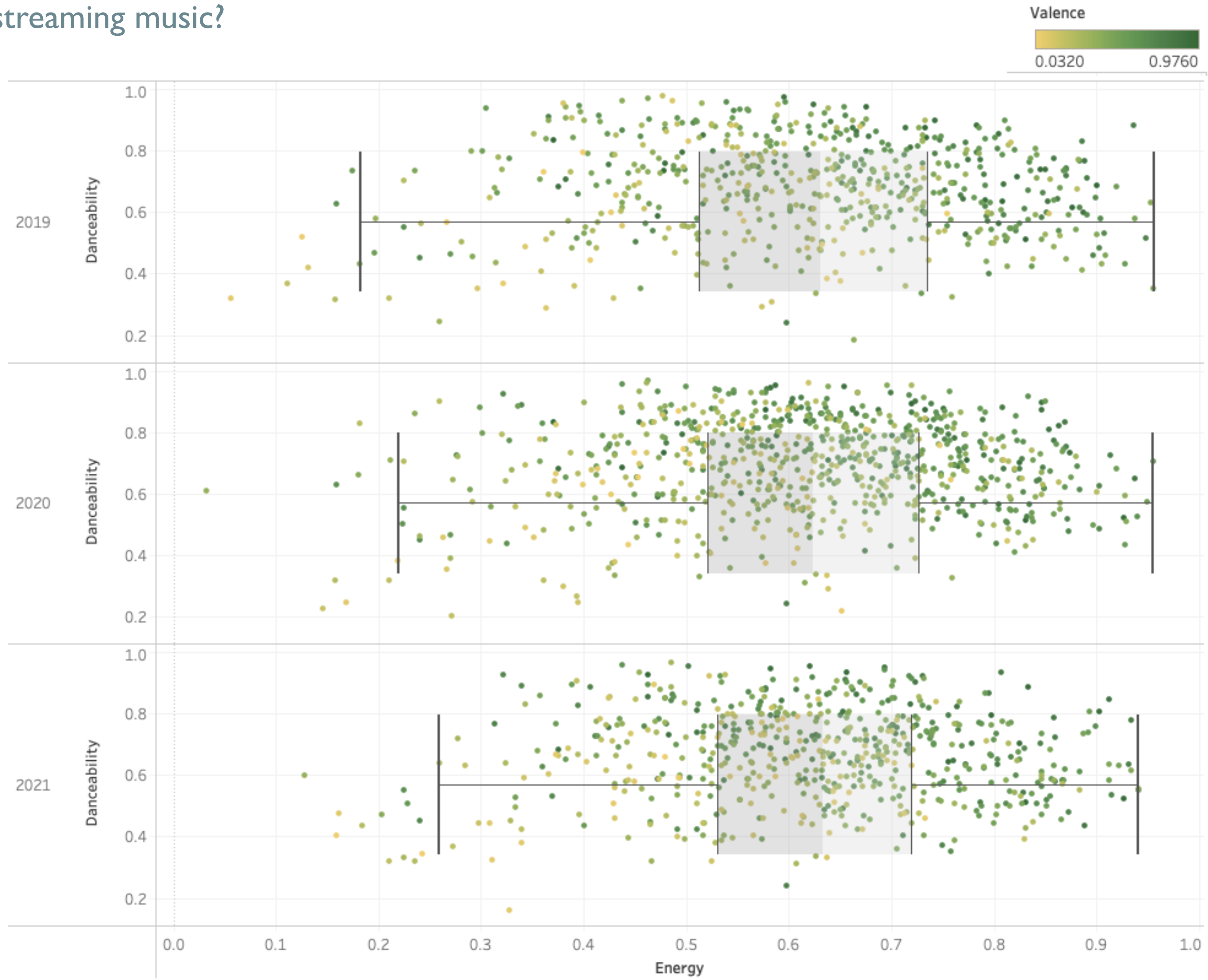
- Main Genres
- Rankings
- **Audio Features**



DEMO What are the changes on streaming music?

Data Visualization Tableau

- Main Genres
- Rankings
- **Audio Features**



DEMO What is the impact of COVID-19 on people's music preferences?

Data Analysis Linear Regression

Danceability

OLS Regression Results

Dep. Variable:	danceability	**R-squared:**	0.729
Model:	OLS	**Adj. R-squared:**	0.708
Method:	Least Squares	**F-statistic:**	35.82
Date:	Thu, 02 Dec 2021	**Prob (F-statistic):**	2.06e-11
Time:	20:38:18	**Log-Likelihood:**	151.76
No. Observations:	44	**AIC:**	-295.5
Df Residuals:	40	**BIC:**	-288.4
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	1.661e-07	1.44e-09	115.548	0.000	1.63e-07	1.69e-07
new_cases	-8.64e-09	3.28e-09	-2.632	0.012	-1.53e-08	-2.01e-09
new_vaccinations	1.071e-09	2.6e-10	4.118	0.000	5.45e-10	1.6e-09
month	-6.085e-10	1.24e-10	-4.888	0.000	-8.6e-10	-3.57e-10
year	0.0003	2.91e-06	115.548	0.000	0.000	0.000
month:year	-1.23e-06	2.52e-07	-4.888	0.000	-1.74e-06	-7.21e-07

Omnibus:	0.703	**Durbin-Watson:**	1.314
Prob(Omnibus):	0.704	**Jarque-Bera (JB):**	0.802
Skew:	-0.251	**Prob(JB):**	0.670
Kurtosis:	2.569	**Cond. No.**	1.45e+23

Energy

OLS Regression Results

Dep. Variable:	energy	**R-squared:**	0.508
Model:	OLS	**Adj. R-squared:**	0.471
Method:	Least Squares	**F-statistic:**	13.78
Date:	Thu, 02 Dec 2021	**Prob (F-statistic):**	2.57e-06
Time:	20:37:58	**Log-Likelihood:**	138.65
No. Observations:	44	**AIC:**	-269.3
Df Residuals:	40	**BIC:**	-262.2
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	1.464e-07	1.94e-09	75.604	0.000	1.43e-07	1.5e-07
new_cases	1.822e-08	4.42e-09	4.119	0.000	9.28e-09	2.72e-08
new_vaccinations	-1.953e-11	3.5e-10	-0.056	0.956	-7.28e-10	6.89e-10
month	6.915e-10	1.68e-10	4.124	0.000	3.53e-10	1.03e-09
year	0.0003	3.91e-06	75.604	0.000	0.000	0.000
month:year	1.398e-06	3.39e-07	4.124	0.000	7.13e-07	2.08e-06

Omnibus:	1.166	**Durbin-Watson:**	1.581
Prob(Omnibus):	0.558	**Jarque-Bera (JB):**	1.066
Skew:	-0.361	**Prob(JB):**	0.587
Kurtosis:	2.753	**Cond. No.**	1.45e+23

Valence

OLS Regression Results

Dep. Variable:	valence	**R-squared:**	0.575
Model:	OLS	**Adj. R-squared:**	0.543
Method:	Least Squares	**F-statistic:**	18.01
Date:	Thu, 02 Dec 2021	**Prob (F-statistic):**	1.49e-07
Time:	20:38:29	**Log-Likelihood:**	124.09
No. Observations:	44	**AIC:**	-240.2
Df Residuals:	40	**BIC:**	-233.0
Df Model:	3		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	1.172e-07	2.7e-09	43.464	0.000	1.12e-07	1.23e-07
new_cases	4.209e-08	6.16e-09	6.837	0.000	2.96e-08	5.45e-08
new_vaccinations	-1.068e-10	4.88e-10	-0.219	0.828	-1.09e-09	8.79e-10
month	-1.112e-10	2.33e-10	-0.476	0.636	-5.83e-10	3.61e-10
year	0.0002	5.45e-06	43.464	0.000	0.000	0.000
month:year	-2.247e-07	4.72e-07	-0.476	0.636	-1.18e-06	7.29e-07

Omnibus:	7.407	**Durbin-Watson:**	0.993
Prob(Omnibus):	0.025	**Jarque-Bera (JB):**	6.313
Skew:	-0.874	**Prob(JB):**	0.0426
Kurtosis:	3.624	**Cond. No.**	1.45e+23

CONCLUSION

The changes on streaming music

- Genre: Significant rise in R&B music. Though Pop and Hip-hop music decreased, they secured the largest amount of tracks on the charts.
- Ranking: R&B and Country music appeared in higher rankings in 2020.
- Audio Features: The Danceability, Energy and Valence of tracks are higher during the pandemic.

The impact of COVID-19 on people's music preferences

- Literature Review: COVID timeline and lockdown policies.

THANK YOU

Billboard -- Weekly Top100 Charts

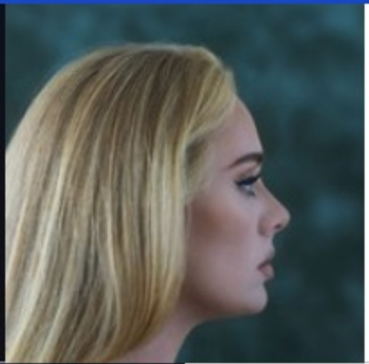
The Weekly Top100 Charts showed the most popular songs across all genres, ranked by radio airplay audience impressions, as measured by Nielsen Music, sales data as compiled by Nielsen Music, and streaming activity data provided by online music sources.

What I did

- Python-Beautiful Soup-Web Scraping: get data from HTML sites
- Output- "BillboardTop100.csv" : with date, ranking, track name, artist name attributes
- Count: 148,000

The Hot 100

WEEK OF NOVEMBER 20, 2021

THIS WEEK			AWARD	LAST WEEK	PEAK POS.	WKS ON CHART
1		**Easy On Me** Adele	★	1	1	5
2		**Stay** The Kid LAROI & Justin Bieber		2	1	18
3		**Industry Baby** Lil Nas X & Jack Harlow		3	1	16
4		**Bad Habits** Ed Sheeran		4	2	20
5		**Shivers**	★	[illegible]	[illegible]	[illegible]

CODE

Billboard

1. Get 148 URLs
2. Get Chart contents: ranking, artist, track

```
import requests
from bs4 import BeautifulSoup as bs
import pandas as pd
#import datetime
from datetime import datetime, date, timedelta


URL = 'https://www.billboard.com/charts/hot-100/'

date_datum = date(2021,11,6)

link_list = []
date_list = []

for year in range(2019,2022):
    if year == 2021:
        week_count_i = (abs(date_datum - date(year, 1, 1)).days)//7
        #print(year,week_count_i)
        for i in range(0, week_count_i):
            link = URL + (str(date_datum - timedelta(7)*i))
            link_list.append(link)
            date_list.append(str(date_datum - timedelta(7)*i))
            #print(link)


    else :
        week_count_j = (abs(date(year,12,31)-date(year, 1, 1)).days)//7
        #print(year,week_count_j)
        for j in range(0, week_count_j):
            link = URL + (str(date(year,12,31) - timedelta(7)*j))
            link_list.append(link)
            date_list.append(str(date(year,12,31) - timedelta(7)*j))
            #print(link)


pd.set_option('display.max_colwidth', None) #show full link
link = pd.DataFrame({'link': link_list, 'date': date_list})

link_chart = pd.concat([link], ignore_index=True).sort_values('date').reset_index(drop=True)

#print(link_chart[0])
```

CODE

Billboard

1. Get 148 URLs
2. Get Chart contents: ranking, artist, track

```
import re


date_list = []
rank_list = []
song_list = []
artist_list = []


for url in link_list:
    req = requests.get(url)
    soup = bs(req.text, 'html.parser')
    #print(url)


    for date in soup.find_all('link', href=re.compile('^/charts/hot-100/')):
        date_text = str(date.get('href')).replace('/charts/hot-100/','')
        date_list.append(date_text)

    for rank in soup.find_all('span','chart-element__rank__number'):
        rank_list.append(str(rank.get_text()))

    for song in soup.find_all('span','chart-element__information__song text--truncate color--primary'):
        song_list.append(str(song.get_text()))

    for artist in soup.find_all('span','chart-element__information__artist text--truncate color--secondary'):
        artist_list.append(str(artist.get_text()))


date_full_list =  [d for d in date_list for i in range(100)]
#print(type(date_full_list[0]))
#print(date_list[0])


chart = pd.DataFrame(
    {'date': date_full_list,
     'rank': rank_list,
     'song': song_list,
     'artist': artist_list})

#print(chart[0:20])
DH_Final_Hot100 = pd.merge(link_chart,chart, on = 'date') #value in both 'date' must be the same == str
#print(DH_Final_Hot100)
```

Spotify – Audio Features

Generate the audio features of the selected songs, including Genre, Beats Per Minute, Energy, Danceability, Loudness, Liveness, Valence, Length, Acousticness, Speechiness, Popularity, Duration.

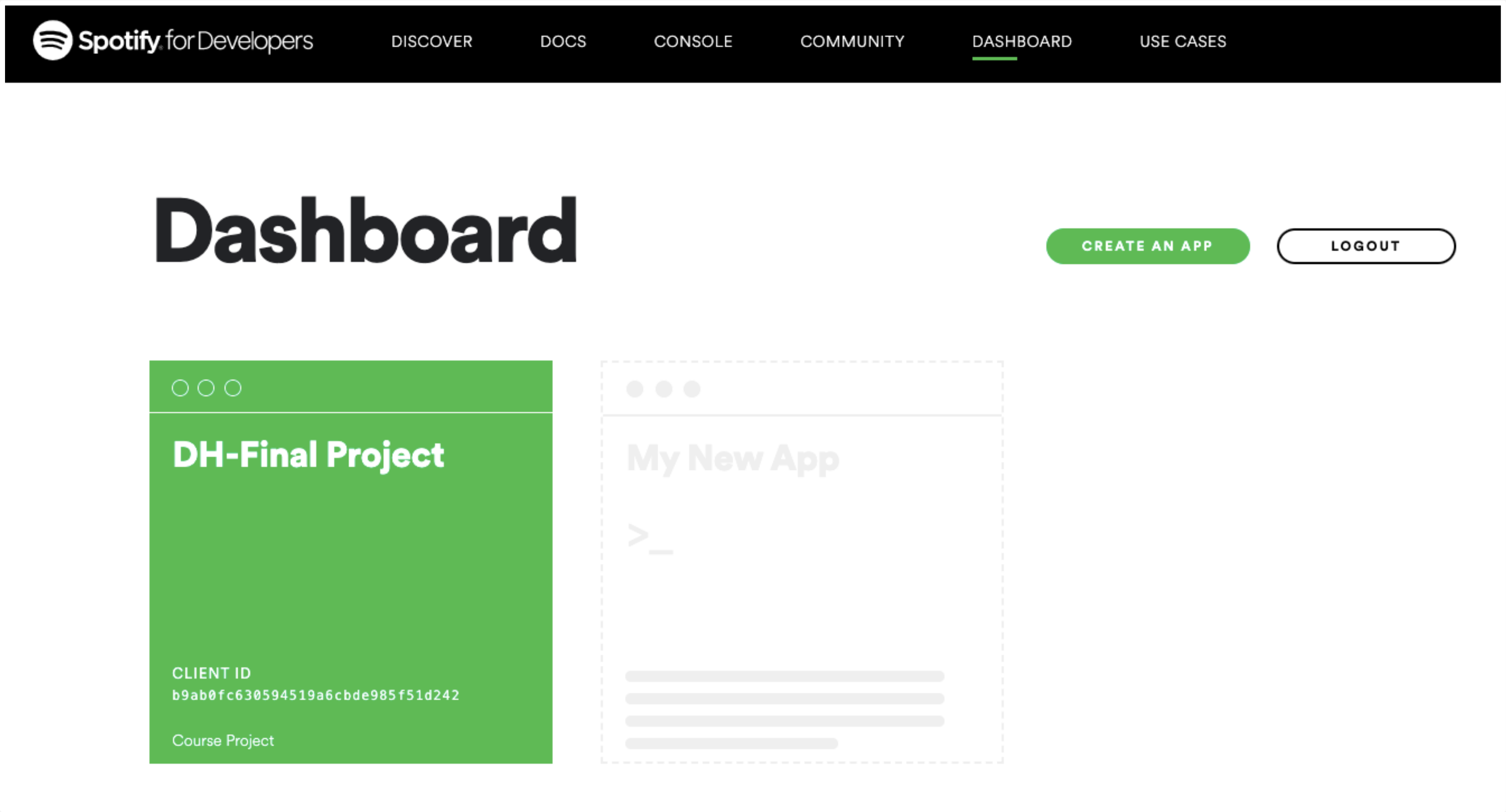
What I did

- Spotipy: a lightweight Python library for the Spotify Web API to get full access to all the music data on Spotify.
- Import "BillboardTop100.csv" to perform query using artist and track name
- Output: track name, artist name, audio features
- Merge with "BillboardTop100.csv" to create complete music dataset

CODE-PREP

Spotify

1. Set authorization flow and get token
2. Import "BillboardTop100.csv". Clean data to prepare for search query
3. Perform search query to get data: track, artist, track id
4. (Check dull value. Add track id manually)
5. Get audio features using track id



CODE

Spotify

1. Set authorization flow and get token
2. Import "BillboardTop100.csv". Clean data to prepare for search query
3. Perform search query to get data: track, artist, track id
4. (Check dull value. Add track id manually)
5. Get audio features using track id

Authorization Code Flow

To support the Authorization Code Flow Spotipy provides a utility method util.prompt_for_user_token that will attempt to authorize the user.

scope = 'user-read-private' token = util.prompt_for_user_token('lauraleetaiwan', scope, client_id='b9ab0fc630594519a6cbde985f51d242',client_secret='59a63526f72e43f5a46b0a06e104ce5e',redirect_uri='http://localhost:8080/') sp = spotipy.Spotify(auth=token)

Client Credentials Flow

Client credentials flow is appropriate for requests that do not require access to a user's private data. Even if you are only making calls that do not require authorization, using this flow yields the benefit of a higher rate limit

client_credentials_manager = SpotifyClientCredentials(client_id='b9ab0fc630594519a6cbde985f51d242',client_secret='59a63526f72e43f5a46b0a06e104ce5e') sp = spotipy.Spotify(client_credentials_manager=client_credentials_manager)

```
##Get 'track id' & 'artist name'
Billboard_Hot100 = pd.read_csv('DH_Final_Hot100.csv')

artist_list=[]
song_list=[]
toBeReplace = ['Featuring', ' X ', ' x ', ' & ', ' + ', ', ']


for artist in Billboard_Hot100['artist']:
    artist_list.append(artist)

for song in Billboard_Hot100['song']:
    song_list.append(song)

for elem in toBeReplace :
    # Check if string is in the main string
    for k,v in enumerate(artist_list):
        if elem in v:
            # Replace the string
            artist_list[k] = v.replace(elem, ' ')
```

CODE

Spotify

1. Set authorization flow and get token
2. Import "BillboardTop100.csv". Clean data to prepare for search query
3. Perform search query to get data: track, artist, track id
4. (Check dull value. Add track id manually)
5. Get audio features using track id

```
track_id_list=[]

error_artist=[]
error_track=[]
error_list=[]

#search from list
for i , j in enumerate(artist_list):
    for m ,n in enumerate(song_list):
        if i<3000 and i== m:
            search_output = sp.search(q='artist:' + j + ' track:' + n, type='track')
            try :
##clean output data
                search_output_1 = pd.DataFrame(search_output['tracks'])
                search_output_2 = search_output_1['items'][0]
                track_id = search_output_2.get('id')
                track_artist = search_output_2['artists'][0]['name']
                track_id_list.append(track_id)
            except:
                error_artist.append(j)
                error_track.append(n)
                    #print('error for', 'artist_index:', j, ' track_index:', n)
```

```
#view error outputs with no duplicate values

for (a,b) in zip(error_artist,error_track):
    error_list.append(a + ' >> ' + b)
    error_list = [x for n, x in enumerate(error_list) if x not in error_list[:n]]
```

CODE

Spotify

1. Set authorization flow and get token
2. Import "BillboardTop100.csv". Clean data to prepare for search query
3. Perform search query to get data: track, artist, track id
4. (Check dull value. Add track id manually)
5. Get audio features using track id

```
##get audio features
audio_feature_list=[]

for track_id in track_id_list:
    audio_feature_output = sp.audio_features(track_id)[0]
    audio_feature_list.append(audio_feature_output)

pd.DataFrame(audio_feature_list)
```

	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	type	
0	0.752	0.488	6	-7.050	1	0.0705	0.29700	0.000009	0.0936	0.533	136.041	audio_features	5p7ujcrUXASCNwRa
1	0.717	0.653	1	-5.634	1	0.0658	0.22900	0.000000	0.1010	0.412	106.966	audio_features	3e9HZxeyfWwjeyPAI
2	0.760	0.479	2	-5.574	1	0.0466	0.55600	0.000000	0.0703	0.913	89.911	audio_features	3KkXRkHbMCARz(
3	0.834	0.730	8	-3.714	1	0.2220	0.00513	0.000000	0.1240	0.446	155.008	audio_features	2xLMifQCjDGFmkI
4	0.579	0.904	5	-2.729	1	0.0618	0.19300	0.000000	0.0640	0.681	82.014	audio_features	1rqqCSm0Qe4I9rU
...	...	...	...	...	...	...	...	...	...	...	...	...	
18752	0.530	0.524	4	-5.751	1	0.0327	0.54400	0.000000	0.1100	0.514	150.050	audio_features	4Ec0q0j8hRTCX4Q

Data Collecting | Data Cleaning | Data Analysis & Viz

Data Analysis

What I will do

- Merge with COVID-19 cases dataset: use logistic regression analysis to measure its influence on music.
- Text Analysis: using tf-idf to analysis lyrics and observe the trend of word choices.

Data Visualization

What I will do

- Tableau: to perform visualization chart. Check if it can embed on website.
- Create a interactive website: use HTML or Wordpress to store and perform the results.

- demo "prototype" 15%
- Guidelines for Demo Presentations
- In 10-15 minutes, your presentation should:
- Introduce your project through summarizing your initial proposal (including your initial research questions, steps for answering those questions, etc…)
- Demo your initial data collection through a brief data biography of your dataset and explanation of the methods you used to collect this data
- Outline the scholarly goals of your dataset, what you plan to do next in the project, and whether you needed to change course from what you outlined in your initial proposal