CS 105 Spotify Genre Predictor Project

Ryan Solanki, Jake Merrill, Krishaan Patel, Vishal Gondi, Ryo Kanda

A glimpse of our dataset…

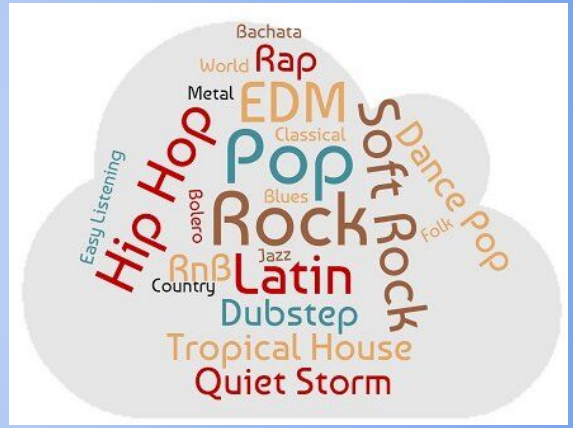
- Contains over 18,000 songs
- Contains categorical variables such as track name and genre
- Contains numerical data about song characteristics

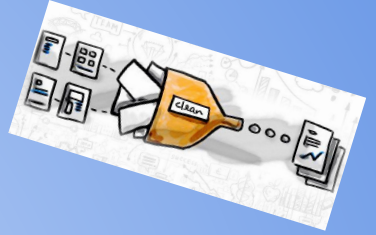
	track_name	track_artist	track_popularity	playlist_genre
2	Pangarap	Barbie's Cradle	41	rock
3	I Feel Alive	Steady Rollin	28	rock
4	Poison	Bell Biv DeVoe	0	r&b
5	Baby It's Cold Outside (feat. Christina Aguilera)	CeeLo Green	41	r&b
6	Dumb Litty	KARD	65	pop
7	Soldier	James TW	70	r&b
8	Satisfy You	Diddy	52	r&b
9	Tender Lover	Babyface	36	r&b
10	Hide Away (feat. Envy Monroe)	Blasterjaxx	42	edm
11	Ti volevo dedicare (feat. J-AX & Boomdabash)	Rocco Hunt	78	r&b
12	Una Vaina Loca	Fuego	1	latin
13	Limestone	Magic City Hippies	58	pop
14	La Mordidita	Ricky Martin	69	latin
15	Changes	2Pac	72	rap
16	Latina (feat. Maluma)	Reykon	74	latin
17	Let Me Entertain You - Remastered 2011	Queen	41	rock
18	Suga Suga	Baby Bash	8	r&b
19	You Don't Know Me - Radio Edit	Jax Jones	14	pop
20	Secrets	The Weeknd	71	r&b
21	Me EnamorÃ©	Jay Wheeler	62	latin
22	Stand Out - From "How to Build a Better Boy"	Sabrina Carpenter	33	pop
23	Laps	Zotiyac	36	rap
24	Get The Funk Out Ma Face	The Brothers Johnson	49	r&b
25	Ooh	SAYMYNAME	39	rap
26	La Vida Es Una (feat. Pitbull)	Lil Jon	32	latin
27	Juke Box Hero	Foreigner	67	pop

M	N	O	P	Q	R	S	T	U	V	W	X
danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms
0.682	0.401	2	-10.068	1	0.0236	0.279	0.0117	0.0887	0.566	97.091	235440
0.303	0.88	9	-4.739	1	0.0442	0.0117	0.00994	0.347	0.404	135.225	373512
0.845	0.652	6	-7.504	0	0.216	0.00432	0.00723	0.489	0.65	111.904	262467
0.425	0.378	5	-5.819	0	0.0341	0.689	0	0.0664	0.405	118.593	243067
0.76	0.887	9	-1.993	1	0.0409	0.037	0	0.138	0.24	130.018	193160
0.496	0.639	6	-6.157	1	0.055	0.28	0	0.0975	0.305	147.764	224720
0.764	0.594	6	-10.05	1	0.185	0.591	0	0.145	0.695	87.261	286441
0.743	0.86	5	-6.346	1	0.0445	0.226	4.22E-04	0.0513	0.687	102.459	259267
0.573	0.746	10	-4.894	1	0.0421	0.0249	0	0.361	0.134	130.001	188000
0.754	0.725	8	-6.058	1	0.0661	0.0104	0	0.192	0.271	120.002	208133
0.794	0.882	1	-5.589	1	0.0361	0.114	7.98E-05	0.062	0.642	117.009	188213
0.706	0.457	4	-9.359	1	0.0383	0.282	0.00849	0.0614	0.723	78.014	209165
0.725	0.903	0	-2.955	1	0.0658	0.0344	0	0.183	0.946	142.006	211680
0.865	0.657	7	-6.722	1	0.076	0.0404	0	0.14	0.316	111.115	268960
0.826	0.713	9	-4.017	0	0.129	0.0748	0	0.331	0.75	102.071	226739
0.521	0.902	11	-4.826	0	0.197	0.0164	0.00105	0.737	0.592	104.708	181693
0.662	0.748	5	-3.041	0	0.268	0.688	8.43E-06	0.0841	0.535	82.331	239027
0.876	0.669	11	-6.054	0	0.138	0.163	0	0.185	0.682	124.007	213947
0.665	0.771	0	-5.779	1	0.0533	0.0155	9.60E-06	0.46	0.764	109.945	265600
0.495	0.357	11	-7.555	1	0.0439	0.968	0.00104	0.118	0.456	179.74	253830
0.414	0.806	7	-2.892	1	0.0665	0.00792	9.33E-05	0.0488	0.77	96.044	208827
0.756	0.453	8	-9.965	1	0.318	0.0039	0	0.632	0.114	140.132	144380
0.88	0.57	2	-13.145	1	0.0821	0.292	0.0388	0.0674	0.791	106.035	147000
0.59	0.965	4	-1.787	1	0.0548	0.0147	0.163	0.12	0.386	148.167	147568
0.742	0.921	9	-3.661	1	0.211	0.00409	0	0.215	0.604	126.04	180591
0.357	0.653	9	-5.554	1	0.0654	0.0828	0	0.0844	0.522	176.647	259800

Project Description…

- Categorize songs into genres based off characteristics such as energy level, loudness, speechiness, acousticness, tempo, etc.
- Analyze the data to find patterns and correlations between music genre and other characteristics
- Create visualizations based off these patterns and correlations
- Create a model that can accurately predict a song's genre based off the above characteristics





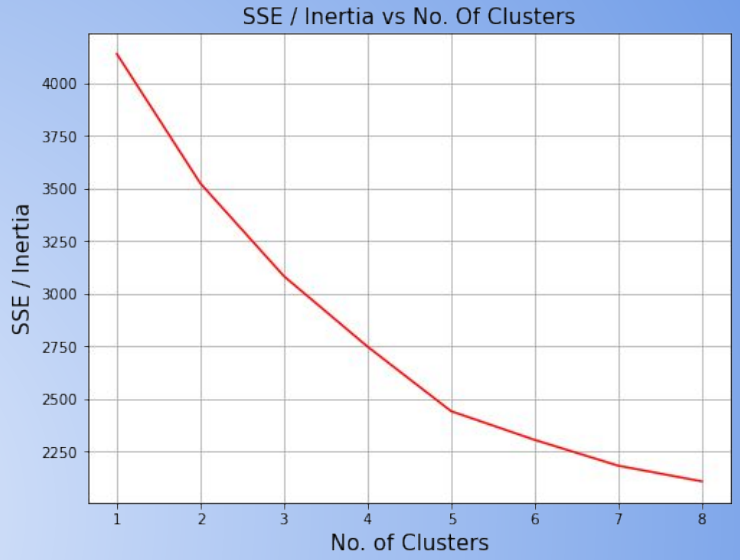
Data Cleaning…

- **Data normalization** - normalized music characteristic values such as energy level, loudness, speechiness, acousticness, tempo, etc. using min-max normalization
- **Data Removal** - removed unnecessary variables from our dataset, such as track IDs and song lyrics
- No null or missing data in our dataset so we did not need to account for this

	track_popularity	danceability	instrumentalness	speechiness	tempo	energy	loudness	duration_ms	valence	acousticness	playlist_genre
0	0.41	0.662692	0.001417	0.342681	0.391225	0.011854	0.588336	0.412526	0.558854	0.281249	3
1	0.28	0.232254	0.025750	0.560561	0.878856	0.010071	0.781738	0.700485	0.390700	0.011793	3
4	0.65	0.751278	0.021852	0.530810	0.885982	0.000000	0.881397	0.324348	0.220469	0.037297	1
8	0.42	0.538898	0.023270	0.530713	0.742441	0.000000	0.776112	0.313586	0.110442	0.025099	0
11	0.58	0.689949	0.018781	0.233684	0.448234	0.008602	0.614067	0.357728	0.721819	0.284273	1

KNN Classifier (Model 1)

- To Begin for KNN we used an elbow plot to find the best K for our data is 5.
- First tried creating a model with all genres and characteristics, but only had an accuracy of about 47%
- In this model we used Euclidean distance with four different genres: edm, pop, rap, rock. (*We dropped latin and r&b*)
- Using K = 5 this model produced an accuracy of .6309 or 63.09



Classification Report:

	precision	recall	f1-score	support
0	0.55	0.62	0.58	422
1	0.51	0.62	0.56	771
2	0.79	0.63	0.70	685
3	0.74	0.65	0.69	712
accuracy			0.63	2590
macro avg	0.65	0.63	0.63	2590
weighted avg	0.65	0.63	0.64	2590

KNN Classifier (Model 2)

- In the second model we also used an elbow plot to determine the best K to use in order to maximize accuracy would be 8.

- In addition to the previous model we also removed the pop genre and slightly changed search variables.

- Using K = 8 this updated model is able to predict the genre with an accuracy of 81.75%.

	precision	recall	f1-score	support
0	0.71	0.69	0.70	423
2	0.86	0.81	0.84	663
3	0.84	0.90	0.87	706
accuracy			0.82	1792
macro avg	0.80	0.80	0.80	1792
weighted avg	0.82	0.82	0.82	1792

Accuracy: 0.8175223214285714

Exploratory Data Analysis…

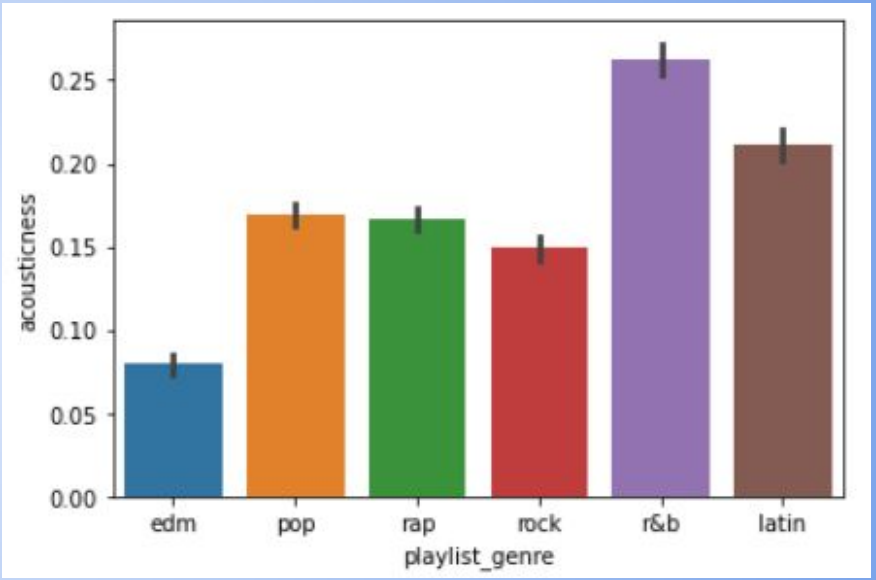
Questions to be answered:

- Which music genre is the most popular?
- Which music genre has the highest acoustic levels?
- Which music genre is the most danceable?
- Which music genre is the most instrumental?
- Which music genre has the highest energy level?

Which genre has the highest acoustic levels?

- Using our normalized data, we created bar plot visualizations of all the main music characteristics using the mean from each genre

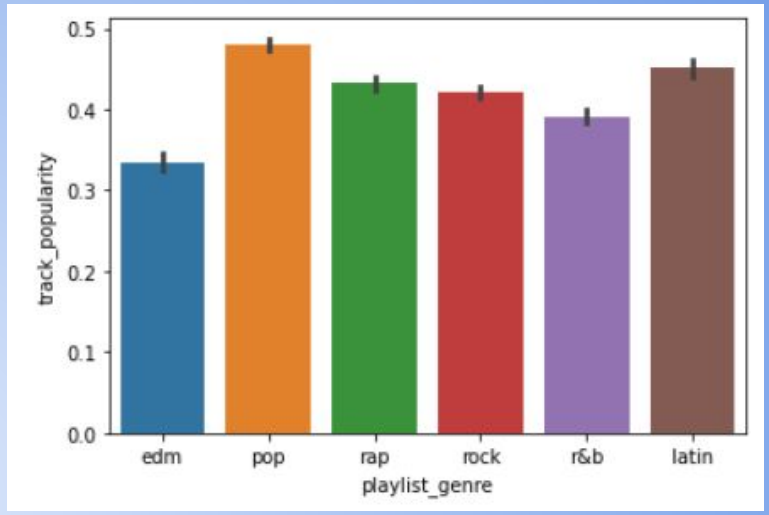
- Based off our analysis, we saw that r&b songs have the highest acoustic levels, whereas EDM has the lowest

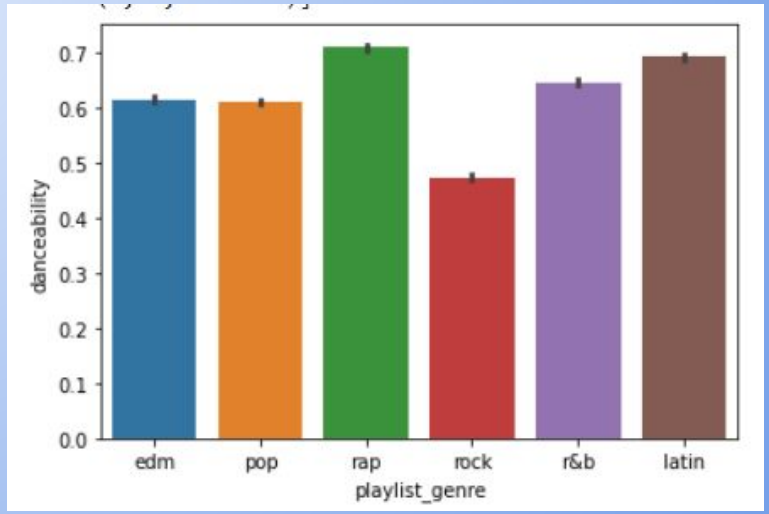


Which genres are the most popular and most danceable?

- Pop is the most popular music genre, with latin, rap, and rock following close behind

- Rap songs have the highest danceability levels, while rock has the lowest

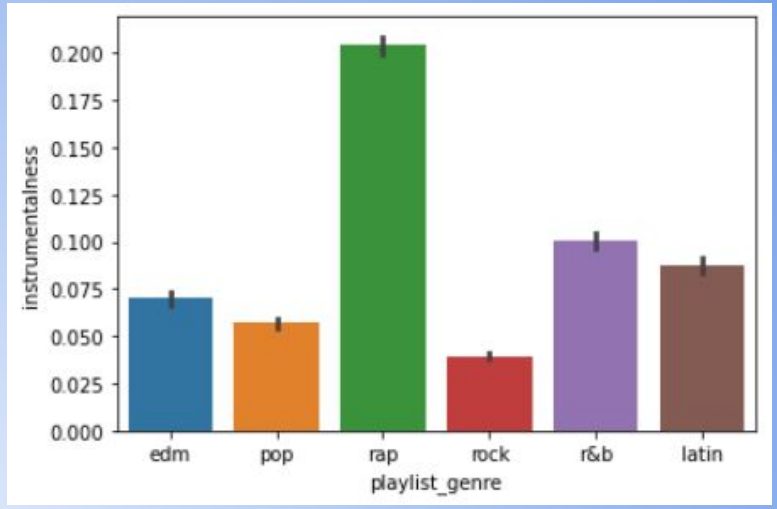


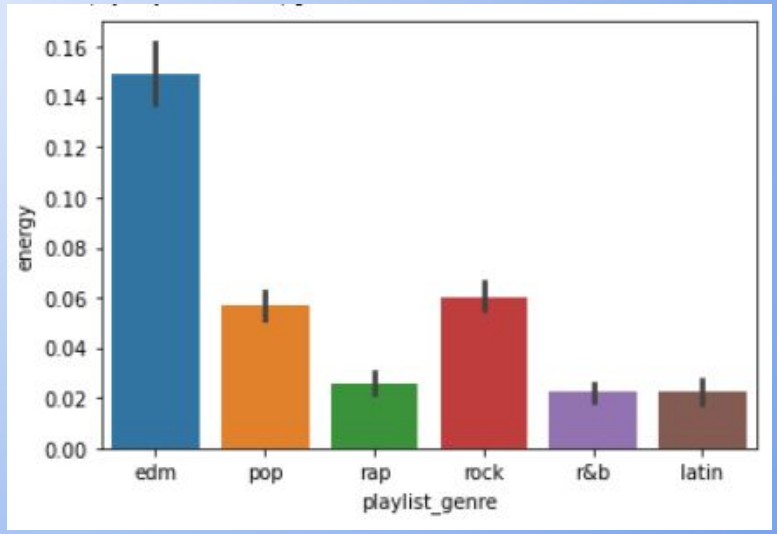


Which genre has the highest instrumentalness and energy level?

- Rap has the highest instrumentalness by far, with rock music having the lowest

- EDM music has the highest energy level, with rap having the lowest





Questions…

1. To get the highest accuracy for KNN model, which genres were dropped from the dataset?

a. Pop, R&B, Latin
b. Rap, EDM, Pop
c. Rock, EDM, R&B
d. EDM, R&B, Latin

2. What was the highest accuracy we were able to achieve?

a. .47
b. .63
c. .81
d. .91

3. Which genre was the most popular in our dataset?

a. Rap
b. Pop
c. Latin
d. EDM