DATE:26/06/23

PROJECT

MUSIC RECOMMENDATION SYSTEM USING K-MEANS & K-NN ALGORITHM

PROBLEM DEFINITION :

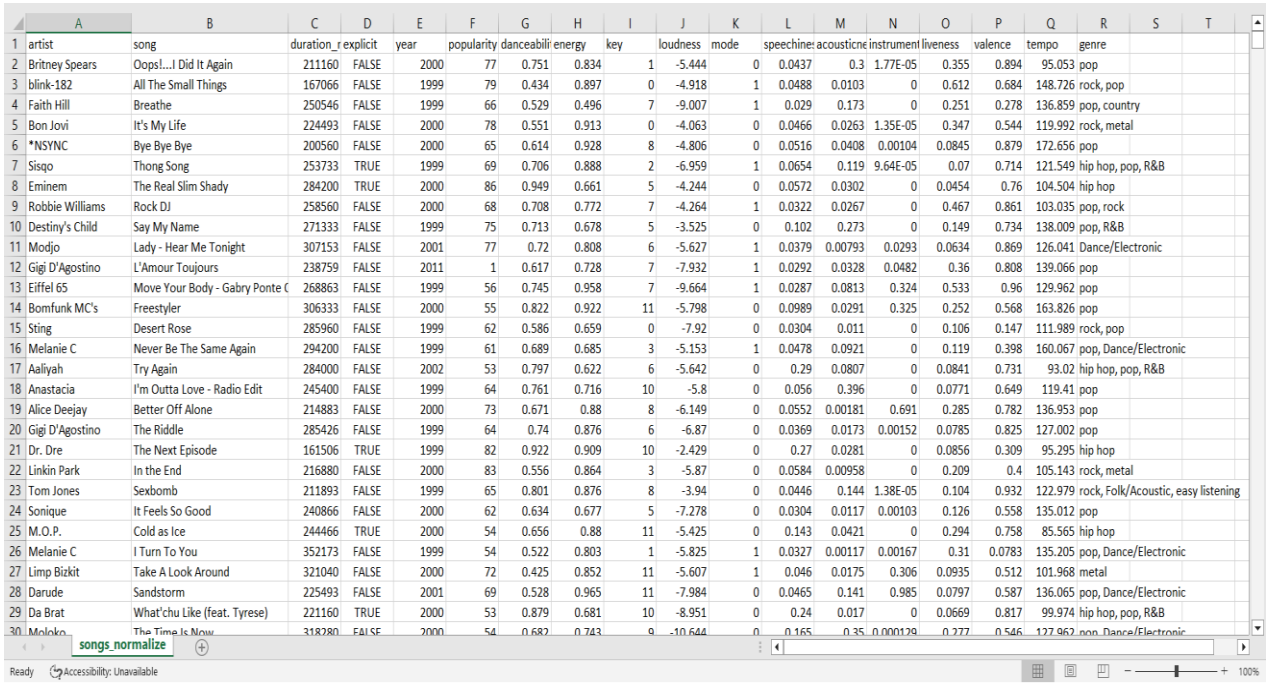
A music recommendation system is to develop a system that can suggest relevant songs or music to users based on their preferences, historical data, and other relevant factors. The goal is to provide personalized recommendations that cater to the user's taste, enhance their music listening experience, and help them discover new songs or artists.

USES OF THE PROJECT :

A music recommendation system enhances the user's music discovery process, improves their listening experience, and benefits music platforms by increasing user engagement and satisfaction.

- Personalized Music Discovery:
- Enhanced User Experience:
- Personalized Playlists and Radio Stations:
- Discovering Niche or Independent Artists:

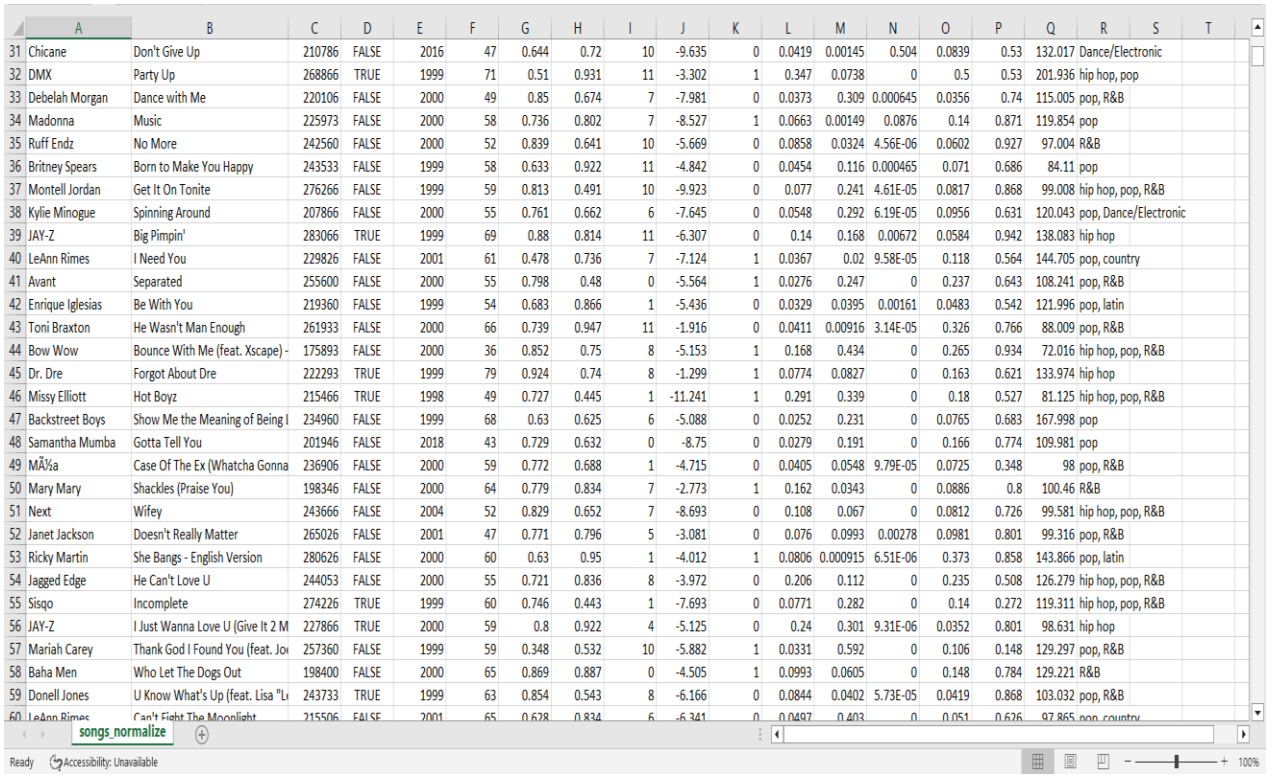
DATASET :



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	artist	song	duration_r	explicit	year	popularity	danceabili	energy	key	loudness	mode	speechines	acousticne	instrument	liveness	valence	tempo	genre
2	Britney Spears	Oops!...I Did It Again	211160	FALSE	2000	77	0.751	0.834	1	-5.444	0	0.0437	0.3	1.77E-05	0.355	0.894	95.053	pop
3	blink-182	All The Small Things	167066	FALSE	1999	79	0.434	0.897	0	-4.918	1	0.0488	0.0103	0	0.612	0.684	148.726	rock, pop
4	Faith Hill	Breathe	250546	FALSE	1999	66	0.529	0.496	7	-9.007	1	0.029	0.173	0	0.251	0.278	136.859	pop, country
5	Bon Jovi	It's My Life	224493	FALSE	2000	78	0.551	0.913	0	-4.063	0	0.0466	0.0263	1.35E-05	0.347	0.544	119.992	rock, metal
6	*NSYNC	Bye Bye Bye	200560	FALSE	2000	65	0.614	0.928	8	-4.806	0	0.0516	0.0408	0.00104	0.0845	0.879	172.656	pop
7	Sisqo	Thong Song	253733	TRUE	1999	69	0.706	0.888	2	-6.959	1	0.0654	0.119	9.64E-05	0.07	0.714	121.549	hip hop, pop, R&B
8	Eminem	The Real Slim Shady	284200	TRUE	2000	86	0.949	0.661	5	-4.244	0	0.0572	0.0302	0	0.0454	0.76	104.504	hip hop
9	Robbie Williams	Rock DJ	258560	FALSE	2000	68	0.708	0.772	7	-4.264	1	0.0322	0.0267	0	0.467	0.861	103.035	pop, rock
10	Destiny's Child	Say My Name	271333	FALSE	1999	75	0.713	0.678	5	-3.525	0	0.102	0.273	0	0.149	0.734	138.009	pop, R&B
11	Modjo	Lady - Hear Me Tonight	307153	FALSE	2001	77	0.72	0.808	6	-5.627	1	0.0379	0.00793	0.0293	0.0634	0.869	126.041	Dance/Electronic
12	Gigi D'Agostino	L'Amour Toujours	238759	FALSE	2011	1	0.617	0.728	7	-7.932	1	0.0292	0.0328	0.0482	0.36	0.808	139.066	pop
13	Eiffel 65	Move Your Body - Gabry Ponte (	268863	FALSE	1999	56	0.745	0.958	7	-9.664	1	0.0287	0.0813	0.324	0.533	0.96	129.962	pop
14	Bomfunk MC's	Freestyler	306333	FALSE	2000	55	0.822	0.922	11	-5.798	0	0.0989	0.0291	0.325	0.252	0.568	163.826	pop
15	Sting	Desert Rose	285960	FALSE	1999	62	0.586	0.659	0	-7.92	0	0.0304	0.011	0	0.106	0.147	111.989	rock, pop
16	Melanie C	Never Be The Same Again	294200	FALSE	1999	61	0.689	0.685	3	-5.153	1	0.0478	0.0921	0	0.119	0.398	160.067	pop, Dance/Electronic
17	Aaliyah	Try Again	284000	FALSE	2002	53	0.797	0.622	6	-5.642	0	0.29	0.0807	0	0.0841	0.731	93.02	hip hop, pop, R&B
18	Anastacia	I'm Outta Love - Radio Edit	245400	FALSE	1999	64	0.761	0.716	10	-5.8	0	0.056	0.396	0	0.0771	0.649	119.41	pop
19	Alice Deejay	Better Off Alone	214883	FALSE	2000	73	0.671	0.88	8	-6.149	0	0.0552	0.00181	0.691	0.285	0.782	136.953	pop
20	Gigi D'Agostino	The Riddle	285426	FALSE	1999	64	0.74	0.876	6	-6.87	0	0.0369	0.0173	0.00152	0.0785	0.825	127.002	pop
21	Dr. Dre	The Next Episode	161506	TRUE	1999	82	0.922	0.909	10	-2.429	0	0.27	0.0281	0	0.0856	0.309	95.295	hip hop
22	Linkin Park	In the End	216880	FALSE	2000	83	0.556	0.864	3	-5.87	0	0.0584	0.00958	0	0.209	0.4	105.143	rock, metal
23	Tom Jones	Sexbomb	211893	FALSE	1999	65	0.801	0.876	8	-3.94	0	0.0446	0.144	1.38E-05	0.104	0.932	122.979	rock, Folk/Acoustic, easy listening
24	Sonique	It Feels So Good	240866	FALSE	2000	62	0.634	0.677	5	-7.278	0	0.0304	0.0117	0.00103	0.126	0.558	135.012	pop
25	M.O.P.	Cold as Ice	244466	TRUE	2000	54	0.656	0.88	11	-5.425	0	0.143	0.0421	0	0.294	0.758	85.565	hip hop
26	Melanie C	I Turn To You	352173	FALSE	1999	54	0.522	0.803	1	-5.825	1	0.0327	0.00117	0.00167	0.31	0.0783	135.205	pop, Dance/Electronic
27	Limp Bizkit	Take A Look Around	321040	FALSE	2000	72	0.425	0.852	11	-5.607	1	0.046	0.0175	0.306	0.0935	0.512	101.968	metal
28	Darude	Sandstorm	225493	FALSE	2001	69	0.528	0.965	11	-7.984	0	0.0465	0.141	0.985	0.0797	0.587	136.065	pop, Dance/Electronic
29	Da Brat	What'chu Like (feat. Tyrese)	221160	TRUE	2000	53	0.879	0.681	10	-8.951	0	0.24	0.017	0	0.0669	0.817	99.974	hip hop, pop, R&B
30	Moloko	The Time Is Now	318280	FALSE	2000	54	0.682	0.743	9	-10.644	0	0.165	0.35	0.000129	0.277	0.546	127.962	pop, Dance/Electronic

songs_normalize

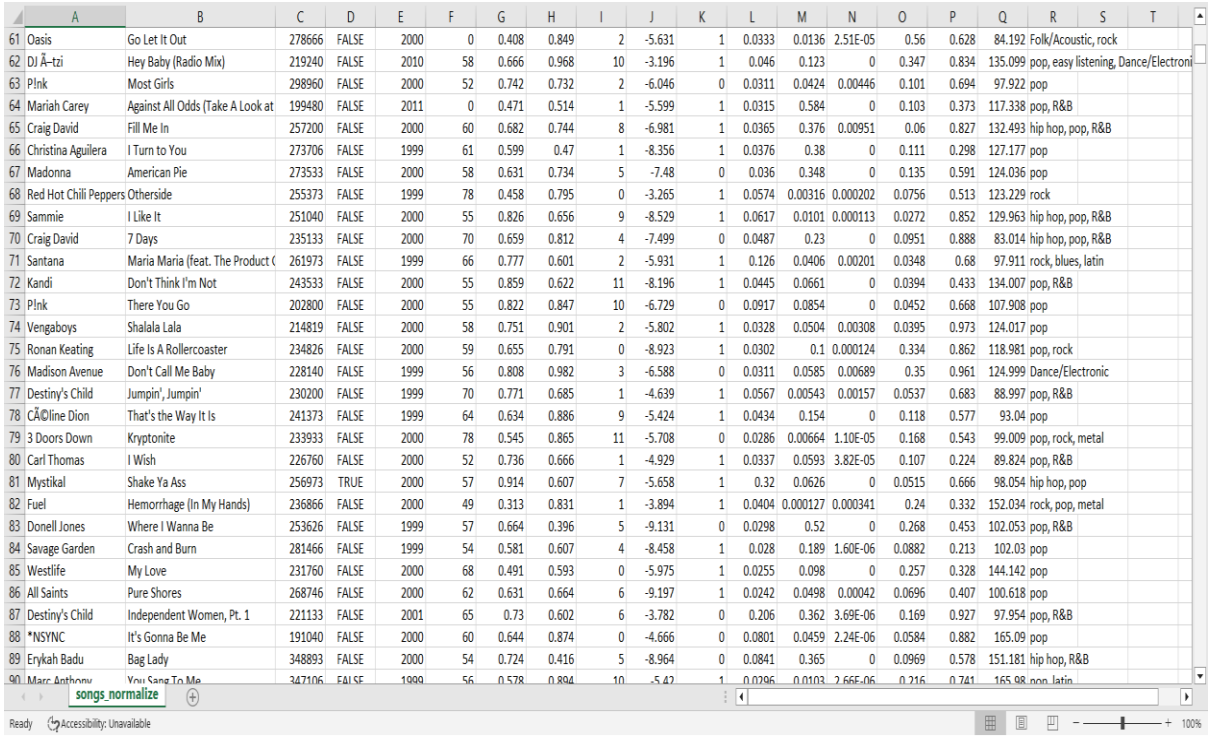
Ready Accessibility: Unavailable 100%



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
31	Chicane	Don't Give Up	210786	FALSE	2016	47	0.644	0.72	10	-9.635	0	0.0419	0.00145	0.504	0.0839	0.53	132.017	Dance/Electronic
32	DMX	Party Up	268866	TRUE	1999	71	0.51	0.931	11	-3.302	1	0.347	0.0738	0	0.5	0.53	201.936	hip hop, pop
33	Debelah Morgan	Dance with Me	220106	FALSE	2000	49	0.85	0.674	7	-7.981	0	0.0373	0.309	0.000645	0.0356	0.74	115.005	pop, R&B
34	Madonna	Music	225973	FALSE	2000	58	0.736	0.802	7	-8.527	1	0.0663	0.00149	0.0876	0.14	0.871	119.854	pop
35	Ruff Endz	No More	242560	FALSE	2000	52	0.839	0.641	10	-5.669	0	0.0858	0.0324	4.56E-06	0.0602	0.927	97.004	R&B
36	Britney Spears	Born to Make You Happy	243533	FALSE	1999	58	0.633	0.922	11	-4.842	0	0.0454	0.116	0.000465	0.071	0.686	84.11	pop
37	Montell Jordan	Get It On Tonite	276266	FALSE	1999	59	0.813	0.491	10	-9.923	0	0.077	0.241	4.61E-05	0.0817	0.868	99.008	hip hop, pop, R&B
38	Kylie Minogue	Spinning Around	207866	FALSE	2000	55	0.761	0.662	6	-7.645	0	0.0548	0.292	6.19E-05	0.0956	0.631	120.043	pop, Dance/Electronic
39	JAY-Z	Big Pimpin'	283066	TRUE	1999	69	0.88	0.814	11	-6.307	0	0.14	0.168	0.00672	0.0584	0.942	138.083	hip hop
40	LeAnn Rimes	I Need You	229826	FALSE	2001	61	0.478	0.736	7	-7.124	1	0.0367	0.02	9.58E-05	0.118	0.564	144.705	pop, country
41	Avant	Separated	255600	FALSE	2000	55	0.798	0.48	0	-5.564	1	0.0276	0.247	0	0.237	0.643	108.241	pop, R&B
42	Enrique Iglesias	Be With You	219360	FALSE	1999	54	0.683	0.866	1	-5.436	0	0.0329	0.0395	0.00161	0.0483	0.542	121.996	pop, latin
43	Toni Braxton	He Wasn't Man Enough	261933	FALSE	2000	66	0.739	0.947	11	-1.916	0	0.0411	0.00916	3.14E-05	0.326	0.766	88.009	pop, R&B
44	Bow Wow	Bounce With Me (feat. Xscape) -	175893	FALSE	2000	36	0.852	0.75	8	-5.153	1	0.168	0.434	0	0.265	0.934	72.016	hip hop, pop, R&B
45	Dr. Dre	Forgot About Dre	222293	TRUE	1999	79	0.924	0.74	8	-1.299	1	0.0774	0.0827	0	0.163	0.621	133.974	hip hop
46	Missy Elliott	Hot Boyz	215466	TRUE	1998	49	0.727	0.445	1	-11.241	1	0.291	0.339	0	0.18	0.527	81.125	hip hop, pop, R&B
47	Backstreet Boys	Show Me the Meaning of Being I	234960	FALSE	1999	68	0.63	0.625	6	-5.088	0	0.0252	0.231	0	0.0765	0.683	167.998	pop
48	Samantha Mumba	Gotta Tell You	201946	FALSE	2018	43	0.729	0.632	0	-8.75	0	0.0279	0.191	0	0.166	0.774	109.981	pop
49	MÃ½a	Case Of The Ex (Whatcha Gonna	236906	FALSE	2000	59	0.772	0.688	1	-4.715	0	0.0405	0.0548	9.79E-05	0.0725	0.348	98	pop, R&B
50	Mary Mary	Shackles (Praise You)	198346	FALSE	2000	64	0.779	0.834	7	-2.773	1	0.162	0.0343	0	0.0886	0.8	100.46	R&B
51	Next	Wifey	243666	FALSE	2004	52	0.829	0.652	7	-8.693	0	0.108	0.067	0	0.0812	0.726	99.581	hip hop, pop, R&B
52	Janet Jackson	Doesn't Really Matter	265026	FALSE	2001	47	0.771	0.796	5	-3.081	0	0.076	0.0993	0.00278	0.0981	0.801	99.316	pop, R&B
53	Ricky Martin	She Bangs - English Version	280626	FALSE	2000	60	0.63	0.95	1	-4.012	1	0.0806	0.000915	6.51E-06	0.373	0.858	143.866	pop, latin
54	Jagged Edge	He Can't Love U	244053	FALSE	2000	55	0.721	0.836	8	-3.972	0	0.206	0.112	0	0.235	0.508	126.279	hip hop, pop, R&B
55	Sisqo	Incomplete	274226	TRUE	1999	60	0.746	0.443	1	-7.693	0	0.0771	0.282	0	0.14	0.272	119.311	hip hop, pop, R&B
56	JAY-Z	I Just Wanna Love U (Give It 2 M	227866	TRUE	2000	59	0.8	0.922	4	-5.125	0	0.24	0.301	9.31E-06	0.0352	0.801	98.631	hip hop
57	Mariah Carey	Thank God I Found You (feat. Jo	257360	FALSE	1999	59	0.348	0.532	10	-5.882	1	0.0331	0.592	0	0.106	0.148	129.297	pop, R&B
58	Baha Men	Who Let The Dogs Out	198400	FALSE	2000	65	0.869	0.887	0	-4.505	1	0.0993	0.0605	0	0.148	0.784	129.221	R&B
59	Donell Jones	U Know What's Up (feat. Lisa "L	243733	TRUE	1999	63	0.854	0.543	8	-6.166	0	0.0844	0.0402	5.73E-05	0.0419	0.868	103.032	pop, R&B
60	LeAnn Rimes	Can't Fight The Moonlight	215506	FALSE	2001	65	0.628	0.834	6	-6.341	0	0.0497	0.403	0	0.051	0.626	97.865	pop, country

songs_normalize

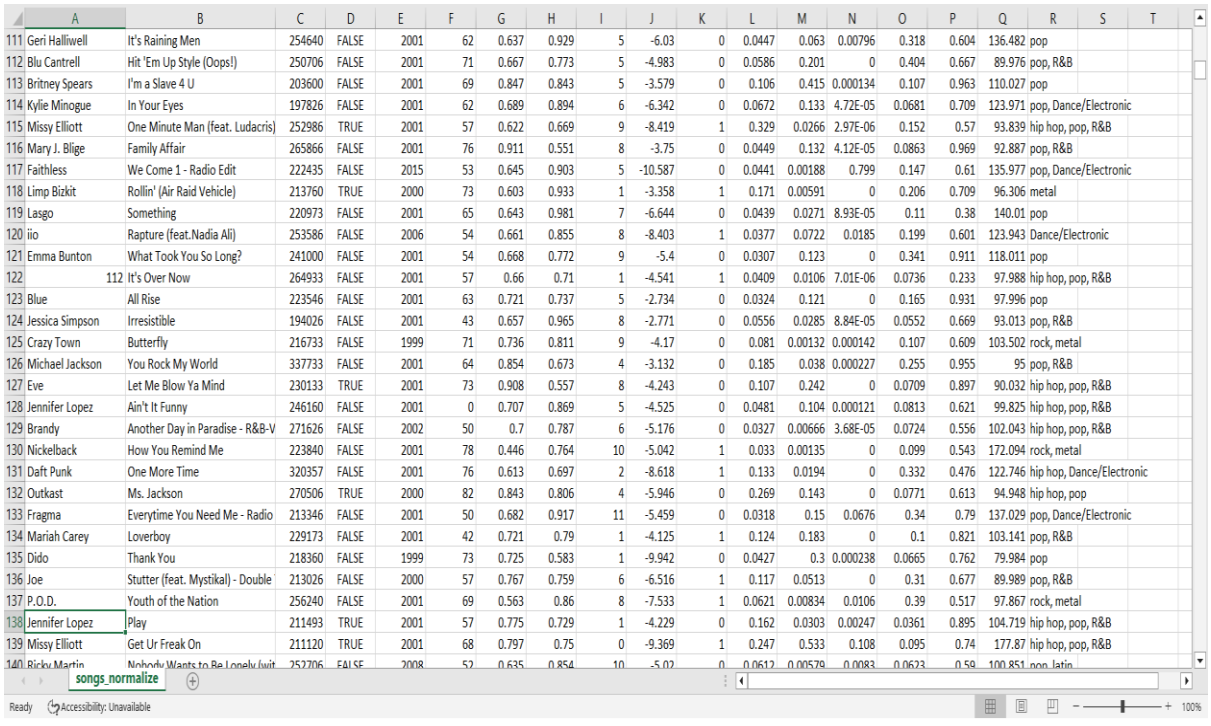
Ready Accessibility: Unavailable 100%



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
61	Oasis	Go Let It Out	278666	FALSE	2000	0	0.408	0.849	2	-5.631	1	0.0333	0.0136	2.51E-05	0.56	0.628	84.192	Folk/Acoustic, rock		
62	DJ Ã–tzi	Hey Baby (Radio Mix)	219240	FALSE	2010	58	0.666	0.968	10	-3.196	1	0.046	0.123	0	0.347	0.834	135.099	pop, easy listening, Dance/Electroni		
63	P!nk	Most Girls	298960	FALSE	2000	52	0.742	0.732	2	-6.046	0	0.0311	0.0424	0.00446	0.101	0.694	97.922	pop		
64	Mariah Carey	Against All Odds (Take A Look at	199480	FALSE	2011	0	0.471	0.514	1	-5.599	1	0.0315	0.584	0	0.103	0.373	117.338	pop, R&B		
65	Craig David	Fill Me In	257200	FALSE	2000	60	0.682	0.744	8	-6.981	1	0.0365	0.376	0.00951	0.06	0.827	132.493	hip hop, pop, R&B		
66	Christina Aguilera	I Turn to You	273706	FALSE	1999	61	0.599	0.47	1	-8.356	1	0.0376	0.38	0	0.111	0.298	127.177	pop		
67	Madonna	American Pie	273533	FALSE	2000	58	0.631	0.734	5	-7.48	0	0.036	0.348	0	0.135	0.591	124.036	pop		
68	Red Hot Chili Peppers	Otherside	255373	FALSE	1999	78	0.458	0.795	0	-3.265	1	0.0574	0.00316	0.000202	0.0756	0.513	123.229	rock		
69	Sammie	I Like It	251040	FALSE	2000	55	0.826	0.656	9	-8.529	1	0.0617	0.0101	0.000113	0.0272	0.852	129.963	hip hop, pop, R&B		
70	Craig David	7 Days	235133	FALSE	2000	70	0.659	0.812	4	-7.499	0	0.0487	0.23	0	0.0951	0.888	83.014	hip hop, pop, R&B		
71	Santana	Maria Maria (feat. The Product (	261973	FALSE	1999	66	0.777	0.601	2	-5.931	1	0.126	0.0406	0.00201	0.0348	0.68	97.911	rock, blues, latin		
72	Kandi	Don't Think I'm Not	243533	FALSE	2000	55	0.859	0.622	11	-8.196	1	0.0445	0.0661	0	0.0394	0.433	134.007	pop, R&B		
73	P!nk	There You Go	202800	FALSE	2000	55	0.822	0.847	10	-6.729	0	0.0917	0.0854	0	0.0452	0.668	107.908	pop		
74	Vengaboys	Shalala Lala	214819	FALSE	2000	58	0.751	0.901	2	-5.802	1	0.0328	0.0504	0.00308	0.0395	0.973	124.017	pop		
75	Ronan Keating	Life Is A Rollercoaster	234826	FALSE	2000	59	0.655	0.791	0	-8.923	1	0.0302	0.1	0.000124	0.334	0.862	118.981	pop, rock		
76	Madison Avenue	Don't Call Me Baby	228140	FALSE	1999	56	0.808	0.982	3	-6.588	0	0.0311	0.0585	0.00689	0.35	0.961	124.999	Dance/Electronic		
77	Destiny's Child	Jumpin', Jumpin'	230200	FALSE	1999	70	0.771	0.685	1	-4.639	1	0.0567	0.00543	0.00157	0.0537	0.683	88.997	pop, R&B		
78	CÃ©line Dion	That's the Way It Is	241373	FALSE	1999	64	0.634	0.886	9	-5.424	1	0.0434	0.154	0	0.118	0.577	93.04	pop		
79	3 Doors Down	Kryptonite	233933	FALSE	2000	78	0.545	0.865	11	-5.708	0	0.0286	0.00664	1.10E-05	0.168	0.543	99.009	pop, rock, metal		
80	Carl Thomas	I Wish	226760	FALSE	2000	52	0.736	0.666	1	-4.929	1	0.0337	0.0593	3.82E-05	0.107	0.224	89.824	pop, R&B		
81	Mystikal	Shake Ya Ass	256973	TRUE	2000	57	0.914	0.607	7	-5.658	1	0.32	0.0626	0	0.0515	0.666	98.054	hip hop, pop		
82	Fuel	Hemorrhage (In My Hands)	236866	FALSE	2000	49	0.313	0.831	1	-3.894	1	0.0404	0.000127	0.000341	0.24	0.332	152.034	rock, pop, metal		
83	Donell Jones	Where I Wanna Be	253626	FALSE	1999	57	0.664	0.396	5	-9.131	0	0.0298	0.52	0	0.268	0.453	102.053	pop, R&B		
84	Savage Garden	Crash and Burn	281466	FALSE	1999	54	0.581	0.607	4	-8.458	1	0.028	0.189	1.60E-06	0.0882	0.213	102.03	pop		
85	Westlife	My Love	231760	FALSE	2000	68	0.491	0.593	0	-5.975	1	0.0255	0.098	0	0.257	0.328	144.142	pop		
86	All Saints	Pure Shores	268746	FALSE	2000	62	0.631	0.664	6	-9.197	1	0.0242	0.0498	0.00042	0.0696	0.407	100.618	pop		
87	Destiny's Child	Independent Women, Pt. 1	221133	FALSE	2001	65	0.73	0.602	6	-3.782	0	0.206	0.362	3.69E-06	0.169	0.927	97.954	pop, R&B		
88	*NSYNC	It's Gonna Be Me	191040	FALSE	2000	60	0.644	0.874	0	-4.666	0	0.0801	0.0459	2.24E-06	0.0584	0.882	165.09	pop		
89	Erykah Badu	Bag Lady	348893	FALSE	2000	54	0.724	0.416	5	-8.964	0	0.0841	0.365	0	0.0969	0.578	151.181	hip hop, R&B		
90	Marc Anthony	You Sang To Me	347106	FALSE	1999	56	0.578	0.894	10	-5.42	1	0.0296	0.0103	2.66E-06	0.216	0.741	165.98	pop, latin		

songs_normalize

Ready Accessibility: Unavailable



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
111	Geri Halliwell	It's Raining Men	254640	FALSE	2001	62	0.637	0.929	5	-6.03	0	0.0447	0.063	0.00796	0.318	0.604	136.482	pop		
112	Blu Cantrell	Hit 'Em Up Style (Oops!)	250706	FALSE	2001	71	0.667	0.773	5	-4.983	0	0.0586	0.201	0	0.404	0.667	89.976	pop, R&B		
113	Britney Spears	I'm a Slave 4 U	203600	FALSE	2001	69	0.847	0.843	5	-3.579	0	0.106	0.415	0.000134	0.107	0.963	110.027	pop		
114	Kylie Minogue	In Your Eyes	197826	FALSE	2001	62	0.689	0.894	6	-6.342	0	0.0672	0.133	4.72E-05	0.0681	0.709	123.971	pop, Dance/Electronic		
115	Missy Elliott	One Minute Man (feat. Ludacris)	252986	TRUE	2001	57	0.622	0.669	9	-8.419	1	0.329	0.0266	2.97E-06	0.152	0.57	93.839	hip hop, pop, R&B		
116	Mary J. Blige	Family Affair	265866	FALSE	2001	76	0.911	0.551	8	-3.75	0	0.0449	0.132	4.12E-05	0.0863	0.969	92.887	pop, R&B		
117	Faithless	We Come 1 - Radio Edit	222435	FALSE	2015	53	0.645	0.903	5	-10.587	0	0.0441	0.00188	0.799	0.147	0.61	135.977	pop, Dance/Electronic		
118	Limp Bizkit	Rollin' (Air Raid Vehicle)	213760	TRUE	2000	73	0.603	0.933	1	-3.358	1	0.171	0.00591	0	0.206	0.709	96.306	metal		
119	Lasgo	Something	220973	FALSE	2001	65	0.643	0.981	7	-6.644	0	0.0439	0.0271	8.93E-05	0.11	0.38	140.01	pop		
120	iio	Rapture (feat.Nadia Ali)	253586	FALSE	2006	54	0.661	0.855	8	-8.403	1	0.0377	0.0722	0.0185	0.199	0.601	123.943	Dance/Electronic		
121	Emma Bunton	What Took You So Long?	241000	FALSE	2001	54	0.668	0.772	9	-5.4	0	0.0307	0.123	0	0.341	0.911	118.011	pop		
122	112	It's Over Now	264933	FALSE	2001	57	0.66	0.71	1	-4.541	1	0.0409	0.0106	7.01E-06	0.0736	0.233	97.988	hip hop, pop, R&B		
123	Blue	All Rise	223546	FALSE	2001	63	0.721	0.737	5	-2.734	0	0.0324	0.121	0	0.165	0.931	97.996	pop		
124	Jessica Simpson	Irresistible	194026	FALSE	2001	43	0.657	0.965	8	-2.771	0	0.0556	0.0285	8.84E-05	0.0552	0.669	93.013	pop, R&B		
125	Crazy Town	Butterfly	216733	FALSE	1999	71	0.736	0.811	9	-4.17	0	0.081	0.00132	0.000142	0.107	0.609	103.502	rock, metal		
126	Michael Jackson	You Rock My World	337733	FALSE	2001	64	0.854	0.673	4	-3.132	0	0.185	0.038	0.000227	0.255	0.955	95	pop, R&B		
127	Eve	Let Me Blow Ya Mind	230133	TRUE	2001	73	0.908	0.557	8	-4.243	0	0.107	0.242	0	0.0709	0.897	90.032	hip hop, pop, R&B		
128	Jennifer Lopez	Ain't It Funny	246160	FALSE	2001	0	0.707	0.869	5	-4.525	0	0.0481	0.104	0.000121	0.0813	0.621	99.825	hip hop, pop, R&B		
129	Brandy	Another Day in Paradise - R&B-V	271626	FALSE	2002	50	0.7	0.787	6	-5.176	0	0.0327	0.00666	3.68E-05	0.0724	0.556	102.043	hip hop, pop, R&B		
130	Nickelback	How You Remind Me	223840	FALSE	2001	78	0.446	0.764	10	-5.042	1	0.033	0.00135	0	0.099	0.543	172.094	rock, metal		
131	Daft Punk	One More Time	320357	FALSE	2001	76	0.613	0.697	2	-8.618	1	0.133	0.0194	0	0.332	0.476	122.746	hip hop, Dance/Electronic		
132	Outkast	Ms. Jackson	270506	TRUE	2000	82	0.843	0.806	4	-5.946	0	0.269	0.143	0	0.0771	0.613	94.948	hip hop, pop		
133	Fragma	Everytime You Need Me - Radio	213346	FALSE	2001	50	0.682	0.917	11	-5.459	0	0.0318	0.15	0.0676	0.34	0.79	137.029	pop, Dance/Electronic		
134	Mariah Carey	Loverboy	229173	FALSE	2001	42	0.721	0.79	1	-4.125	1	0.124	0.183	0	0.1	0.821	103.141	pop, R&B		
135	Dido	Thank You	218360	FALSE	1999	73	0.725	0.583	1	-9.942	0	0.0427	0.3	0.000238	0.0665	0.762	79.984	pop		
136	Joe	Stutter (feat. Mystikal) - Double	213026	FALSE	2000	57	0.767	0.759	6	-6.516	1	0.117	0.0513	0	0.31	0.677	89.989	pop, R&B		
137	P.O.D.	Youth of the Nation	256240	FALSE	2001	69	0.563	0.86	8	-7.533	1	0.0621	0.00834	0.0106	0.39	0.517	97.867	rock, metal		
138	Jennifer Lopez	Play	211493	TRUE	2001	57	0.775	0.729	1	-4.229	0	0.162	0.0303	0.00247	0.0361	0.895	104.719	hip hop, pop, R&B		
139	Missy Elliott	Get Ur Freak On	211120	TRUE	2001	68	0.797	0.75	0	-9.369	1	0.247	0.533	0.108	0.095	0.74	177.87	hip hop, pop, R&B		
140	Ricky Martin	Nobody Wants to Be Lonely (wit	252706	FALSE	2008	52	0.635	0.854	10	-5.02	0	0.0612	0.00579	0.0083	0.0623	0.59	100.851	pop, latin		

songs_normalize

Ready Accessibility: Unavailable

SOURCE CODE :

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sb

from tqdm import tqdm


from sklearn.metrics.pairwise import cosine_similarity

from sklearn.feature_extraction.text import CountVectorizer

from sklearn.manifold import TSNE

from sklearn.preprocessing import StandardScaler

from sklearn.pipeline import Pipeline


import warnings

warnings.filterwarnings('ignore')


music_data = pd.read_csv('C:/Users/anand/Desktop/ML/songs_normalize.csv')

music_data.head()


music_data.isnull().sum()


music_data.info()


music_data.shape
```

```
music_data.isnull().sum().plot.bar()
plt.show()

music_data.select_dtypes(np.number)

music_data["explicit"] = music_data["explicit"].astype(int)

music_data.head()

visual_data = music_data.drop(columns=['song', 'artist', 'year', 'genre'])

plt.figure(figsize=(20, 20))
for i in tqdm(np.arange(1, len(visual_data.columns))):
    plt.subplot(9, 2, i)
    sb.barplot(x=music_data.year,y=visual_data[visual_data.columns[i]])
    plt.xticks(rotation=45);
plt.show()

plt.subplots(figsize=(12, 8))
sb.heatmap(visual_data.corr(), annot=True, square=True)
plt.show()

from sklearn.preprocessing import OneHotEncoder
unique_genres = set()
for genre_list in music_data["genre"]:
```

```
    genres = genre_list.split(",")
    for genre in genres:
        unique_genres.add(genre)

# Create a one-hot encoding for the genre column
encoder = OneHotEncoder()
encoder.fit([[genre] for genre in unique_genres])

# Encode the genre data
encoded_genres = []
for genres in music_data["genre"]:
    genres = genres.split(",")
    one_hot = [0 if genre not in genres else 1 for genre in unique_genres]
    encoded_genres.append(one_hot)

import os
import seaborn as sns
import plotly.express as px
import matplotlib.pyplot as plt
%matplotlib inline

from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from sklearn.metrics import euclidean_distances
from scipy.spatial.distance import cdist
```

```
import warnings

warnings.filterwarnings("ignore")


def normalize_column(col):

    max_d = music_data[col].max()

    min_d = music_data[col].min()

    music_data[col] = (music_data[col] - min_d)/(max_d - min_d)


num_types = ['int16', 'int32', 'int64', 'float16', 'float32', 'float64']

num = music_data.select_dtypes(include=num_types)


for col in num.columns:

    if col != 'year':

        normalize_column(col)


music_data.select_dtypes(np.number).drop(columns = ['year']).plot(kind='box'
,figsize=(20, 20) ,fontsize=10)


model = TSNE(n_components = 2, random_state = 0)

music_data_modified =
music_data.select_dtypes(np.number).drop(columns=['year'])

tsne_data = model.fit_transform(music_data_modified)

plt.style.use('seaborn')

plt.figure(figsize = (7, 7))

plt.scatter(tsne_data[:,0], tsne_data[:,1],  marker= '*')
```

```
plt.show()


# Create a DataFrame from encoded genres

encoded_genres_df = pd.DataFrame(encoded_genres,
columns=list(unique_genres))


# Concatenate the encoded genres DataFrame with the original dataset

music_data = pd.concat([music_data, encoded_genres_df], axis=1)


# View the dataset with the encoded genres

print(music_data.head())


from sklearn.cluster import KMeans


km = KMeans(n_clusters=10)

cat = km.fit_predict(num)

music_data['cat'] = cat

normalize_column('cat')


music_data.cat[:10]


cluster_pipeline = Pipeline([('scaler', StandardScaler()), ('kmeans',
KMeans(n_clusters=10))])

X = music_data.select_dtypes(np.number)

cluster_pipeline.fit(X)

music_data['cluster'] = cluster_pipeline.predict(X)
```

```
import plotly.express as px


tsne_pipeline = Pipeline([('scaler', StandardScaler()), ('tsne',
TSNE(n_components=2, verbose=1))])

genre_embedding = tsne_pipeline.fit_transform(X)

projection = pd.DataFrame(columns=['x', 'y'], data=genre_embedding)

projection['genres'] = music_data['genre']

projection['cluster'] = music_data['cluster']


clu = px.scatter(projection, x='x', y='y', color='cluster', hover_data=['x', 'y', 'genres'])


clu

clu.show()




from sklearn.neighbors import KNeighborsClassifier

from sklearn.model_selection import train_test_split

X = music_data.select_dtypes(np.number).drop(columns =
['cat','cluster','year']).copy()

y = music_data['cluster']




X_train, X_rem, y_train, y_rem = train_test_split(X,y, train_size=0.8,
random_state=0)


X_valid, X_test, y_valid, y_test = train_test_split(X_rem,y_rem, test_size=0.5)
```

```
print(X_train.shape), print(y_train.shape)
print(X_valid.shape), print(y_valid.shape)
print(X_test.shape), print(y_test.shape)

knn1= KNeighborsClassifier(metric='cosine', algorithm='brute', n_neighbors=1)
knn5= KNeighborsClassifier(metric='cosine', algorithm='brute', n_neighbors=5)
knn10= KNeighborsClassifier(metric='cosine', algorithm='brute', n_neighbors=10)

knn5.fit(X_train, y_train)
knn1.fit(X_train, y_train)
knn10.fit(X_train, y_train)

knn5.fit(X_valid, y_valid)
knn1.fit(X_valid, y_valid)
knn10.fit(X_valid, y_valid)

knn5.fit(X_train, y_train)
knn1.fit(X_train, y_train)
knn10.fit(X_train, y_train)

y_pred_5 = knn5.predict(X_valid)
y_pred_1 = knn1.predict(X_valid)
y_pred_10 = knn1.predict(X_valid)
```

```
from sklearn.metrics import accuracy_score

print("Accuracy with k=5", accuracy_score(y_valid, y_pred_5)*100)

print("Accuracy with k=1", accuracy_score(y_valid, y_pred_1)*100)

print("Accuracy with k=10", accuracy_score(y_valid, y_pred_10)*100)


from sklearn.metrics import classification_report, confusion_matrix

print(confusion_matrix(y_valid, y_pred_1))

print(confusion_matrix(y_valid, y_pred_5))

print(confusion_matrix(y_valid, y_pred_10))


print(classification_report(y_valid, y_pred_1))

print(classification_report(y_valid, y_pred_5))

print(classification_report(y_valid, y_pred_10))


print(classification_report(y_valid, y_pred_1))

print(classification_report(y_valid, y_pred_5))

print(classification_report(y_valid, y_pred_10))


cmap = sb.cubehelix_palette(as_cmap=True)

plt.figure(figsize = (15,5))

plt.subplot(1,2,1)

plt.scatter(tsne_data_X_valid[:,0], tsne_data_X_valid[:,1], c=y_pred_5, marker= '*',
s=100, cmap=cmap)

plt.title("Predicted values with k=5", fontsize=20)

plt.subplot(1,2,2)
```

```
plt.scatter(tsne_data_X_valid[:,0], tsne_data_X_valid[:,1], c=y_pred_1, marker= '*', s=100, cmap=cmap)

plt.title("Predicted values with k=1", fontsize=20)

plt.show()

plt.subplot(1,2,2)

plt.scatter(tsne_data_X_valid[:,0], tsne_data_X_valid[:,1], c=y_pred_10, marker= '*', s=100, cmap=cmap)

plt.title("Predicted values with k=10", fontsize=20)

plt.show()


from fuzzywuzzy import process

X_test


recommendation_set = music_data.merge(X_test, how = 'inner' ,indicator=False)

recommendation_set


def recommender(song_name, data,model):

    idx=process.extractOne(song_name, recommendation_set['song'])[2]

    print('Song Selected:-',recommendation_set['song'][idx], 'Index: ',idx)

    print('Searching for recommendations.....')

    requiredSongs = recommendation_set.select_dtypes(np.number).drop(columns = ['cat','cluster','year']).copy()

    distances, indices = model.kneighbors(requiredSongs.iloc[idx].values.reshape(1,-1))

    for i in indices:

        print(music_data['song'][i] + "       " + music_data['artist'][i])
```

```
def get_song_info(row_number):
    song_info = recommendation_set.loc[row_number, ["song", "artist"]]
    return song_info

# Get user input for song name and artist name
user_song = input("Enter the song name: ")
user_artist = input("Enter the artist name: ")

# Find the row number of the song in the recommendation_set
matching_rows = recommendation_set[(recommendation_set["song"] == user_song)
& (recommendation_set["artist"] == user_artist)]

if len(matching_rows) == 0:
    print("Song not found.")
else:
    row_number = matching_rows.index[0]

    # Get the song info using the row number
    song_info = get_song_info(row_number)

    print("Song name: ", song_info["song"])
    print("Artist name: ", song_info["artist"])

    # Use the song name for recommendation
    song_name = song_info["song"]
    recommender(song_name, X_test, knn5)
```

RESULT:

#DISPLAYING THE HEAD OF THE DATASET

	artist	song	duration_ms	explicit	year	popularity	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	genre
0	Britney Spears	Oops!...I Did It Again	211160	False	2000	77	0.751	0.834	1	-5.444	0	0.0437	0.3000	0.000018	0.3550	0.894	95.053	pop
1	blink-182	All The Small Things	167066	False	1999	79	0.434	0.897	0	-4.918	1	0.0488	0.0103	0.000000	0.6120	0.684	148.726	rock, pop
2	Faith Hill	Breathe	250546	False	1999	66	0.529	0.496	7	-9.007	1	0.0290	0.1730	0.000000	0.2510	0.278	136.859	pop, country
3	Bon Jovi	It's My Life	224493	False	2000	78	0.551	0.913	0	-4.063	0	0.0466	0.0263	0.000013	0.3470	0.544	119.992	rock, metal
4	*NSYNC	Bye Bye Bye	200560	False	2000	65	0.614	0.928	8	-4.806	0	0.0516	0.0408	0.001040	0.0845	0.879	172.656	pop

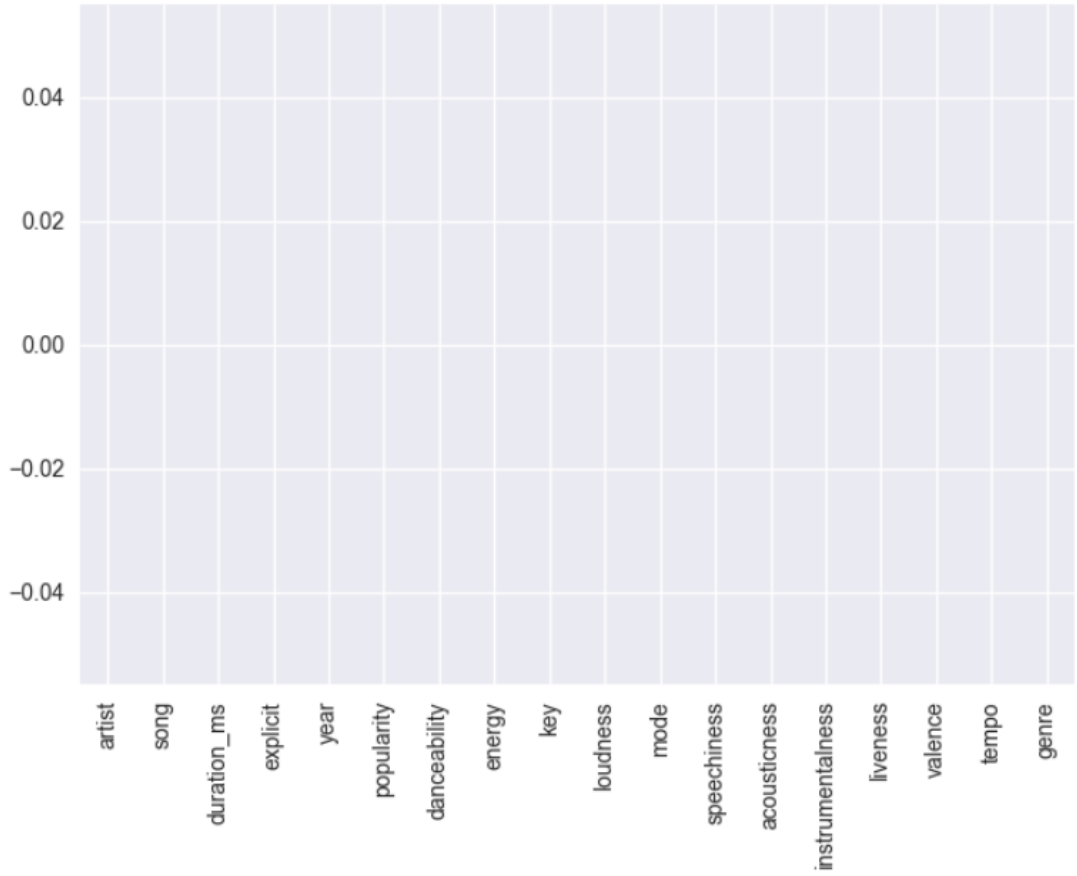
#CHECKING FOR NULL VALUES AND REMOVING

```
artist              0
song                0
duration_ms         0
explicit            0
year                0
popularity          0
danceability        0
energy              0
key                 0
loudness            0
mode                0
speechiness         0
acousticness        0
instrumentalness    0
liveness            0
valence             0
tempo               0
genre               0
dtype: int64
```

#VIEWING THE DATATYPE OF THE DATASET

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 18 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   artist            2000 non-null   object
 1   song              2000 non-null   object
 2   duration_ms       2000 non-null   int64
 3   explicit          2000 non-null   bool
 4   year              2000 non-null   int64
 5   popularity        2000 non-null   int64
 6   danceability      2000 non-null   float64
 7   energy            2000 non-null   float64
 8   key               2000 non-null   int64
 9   loudness          2000 non-null   float64
 10  mode              2000 non-null   int64
 11  speechiness       2000 non-null   float64
 12  acousticness      2000 non-null   float64
 13  instrumentalness  2000 non-null   float64
 14  liveness          2000 non-null   float64
 15  valence           2000 non-null   float64
 16  tempo             2000 non-null   float64
 17  genre             2000 non-null   object
dtypes: bool(1), float64(9), int64(5), object(3)
memory usage: 267.7+ KB
```

#SHOWING THE NULL VALUE



#SELECTING THE COLOUMNS WHICH HAVE NUMERICAL VALUES

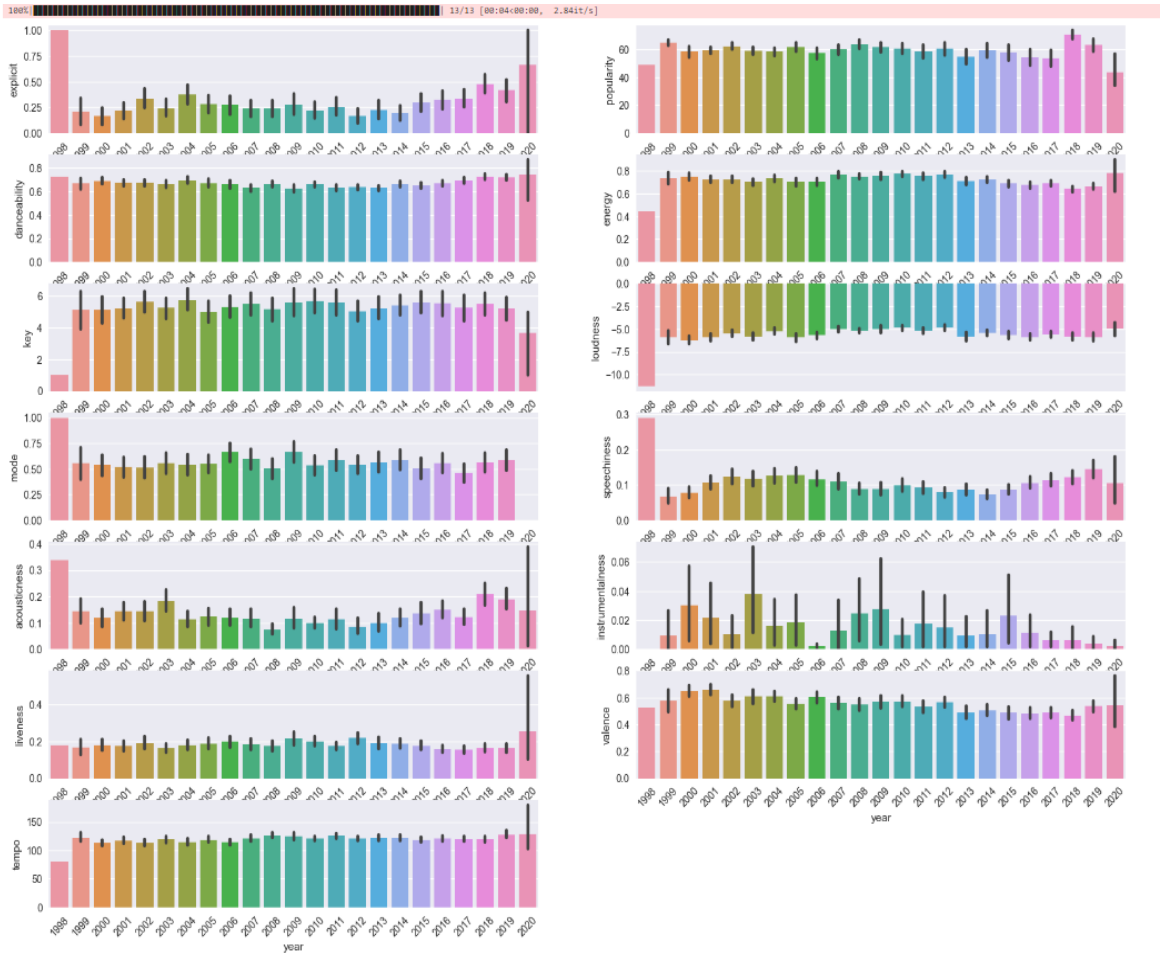
	duration_ms	year	popularity	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo
0	211160	2000	77	0.751	0.834	1	-5.444	0	0.0437	0.3000	0.000018	0.3550	0.894	95.053
1	167066	1999	79	0.434	0.897	0	-4.918	1	0.0488	0.0103	0.000000	0.6120	0.684	148.726
2	250546	1999	66	0.529	0.496	7	-9.007	1	0.0290	0.1730	0.000000	0.2510	0.278	136.859
3	224493	2000	78	0.551	0.913	0	-4.063	0	0.0466	0.0263	0.000013	0.3470	0.544	119.992
4	200560	2000	65	0.614	0.928	8	-4.806	0	0.0516	0.0408	0.001040	0.0845	0.879	172.656
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1995	181026	2019	79	0.842	0.734	1	-5.065	0	0.0588	0.0427	0.000000	0.1060	0.952	137.958
1996	178426	2019	78	0.552	0.702	9	-5.707	1	0.1570	0.1170	0.000021	0.1050	0.564	169.994
1997	200593	2019	69	0.847	0.678	9	-8.635	1	0.1090	0.0669	0.000000	0.2740	0.811	97.984
1998	171029	2019	75	0.741	0.520	8	-7.513	1	0.0656	0.4500	0.000002	0.2220	0.347	102.998
1999	215280	2019	85	0.695	0.762	0	-3.497	1	0.0395	0.1920	0.002440	0.0863	0.553	120.042

2000 rows × 14 columns

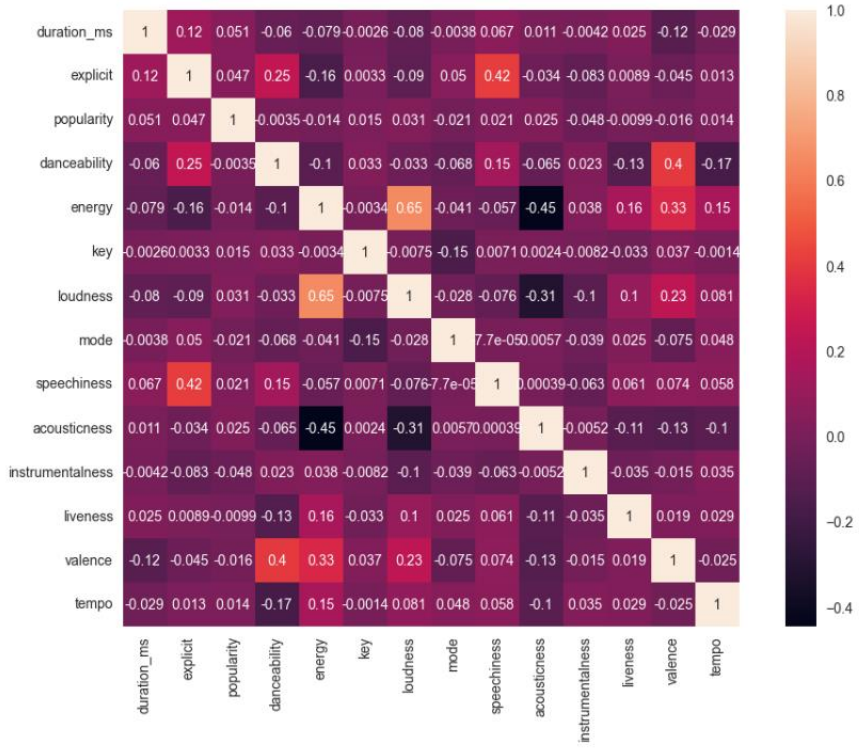
#CHANGING THE EXPLICIT VALUE INTO INT VALUES

	artist	song	duration_ms	explicit	year	popularity	danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	genre
0	Britney Spears	Oops!...I Did It Again	211160	0	2000	77	0.751	0.834	1	-5.444	0	0.0437	0.3000	0.000018	0.3550	0.894	95.053	pop
1	blink-182	All The Small Things	167066	0	1999	79	0.434	0.897	0	-4.918	1	0.0488	0.0103	0.000000	0.6120	0.684	148.726	rock, pop
2	Faith Hill	Breathe	250546	0	1999	66	0.529	0.496	7	-9.007	1	0.0290	0.1730	0.000000	0.2510	0.278	136.859	pop, country
3	Bon Jovi	It's My Life	224493	0	2000	78	0.551	0.913	0	-4.063	0	0.0466	0.0263	0.000013	0.3470	0.544	119.992	rock, metal
4	*NSYNC	Bye Bye Bye	200560	0	2000	65	0.614	0.928	8	-4.806	0	0.0516	0.0408	0.001040	0.0845	0.879	172.656	pop

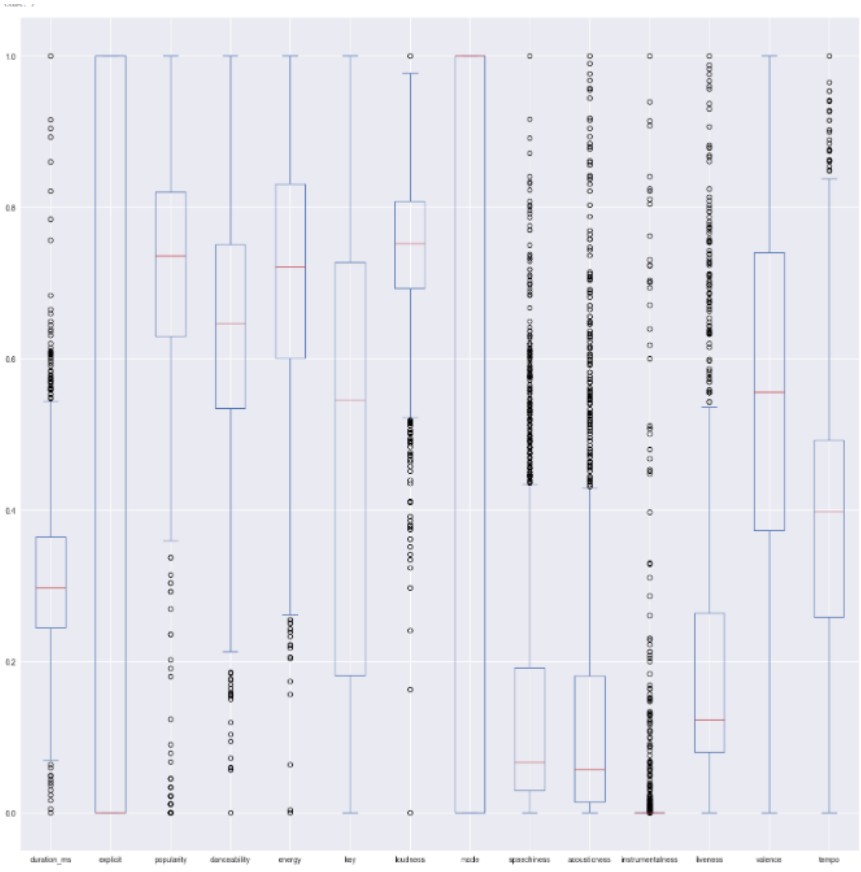
#DATA VISUALIZATION FOR THE DATA IN THE DATASET



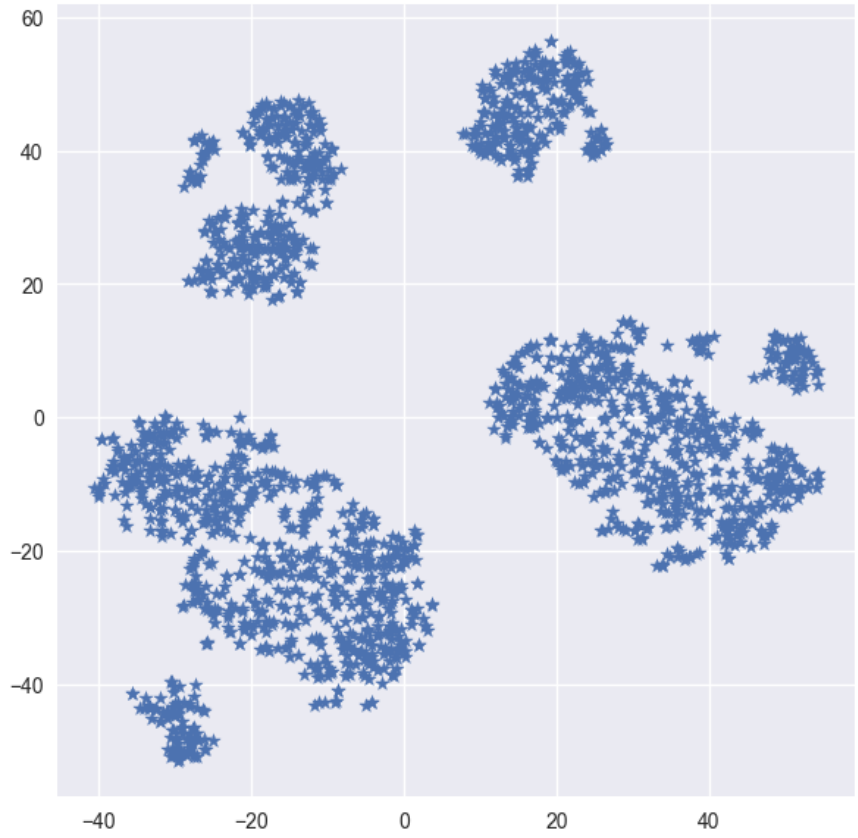
#VISUALISING THE DATA AS HEATMAP



#PLOTTING BARPLOT FOR ALL THE COLUMNS BASED ON THE YEAR



#DATA VISUALISATION USING t-SNE



#ENCODING INTO A UNIQUE GENRE

	artist	song	duration_ms	explicit	year	popularity	danceability	energy	key	loudness	...	R&B	Dance/Electronic	Folk/Acoustic	latin	rock	jazz	country	country	easy listening	blues
0	Britney Spears	Oops!...I Did It Again	0.264478	0.0	2000	0.865169	0.735225	0.825230	0.090909	0.744639	...	0	0	0	0	0	0	0	0	0	0
1	blink-182	All The Small Things	0.145673	0.0	1999	0.887640	0.360520	0.891961	0.000000	0.770630	...	0	0	0	0	0	0	0	0	0	0
2	Faith Hill	Breathe	0.370598	0.0	1999	0.741573	0.472813	0.467217	0.636364	0.568584	...	0	0	0	0	0	0	0	1	0	0
3	Bon Jovi	It's My Life	0.300402	0.0	2000	0.876404	0.498818	0.908908	0.000000	0.812877	...	0	0	0	0	0	0	0	0	0	0
4	*NSYNC	Bye Bye Bye	0.235918	0.0	2000	0.730337	0.573286	0.924796	0.727273	0.776164	...	0	0	0	0	0	0	0	0	0	0

#VISUALISING THE CLUSTER DATA IN t-SNE

```
[t-SNE] Computing 91 nearest neighbors...
[t-SNE] Indexed 2000 samples in 0.001s...
[t-SNE] Computed neighbors for 2000 samples in 0.162s...
[t-SNE] Computed conditional probabilities for sample 1000 / 2000
[t-SNE] Computed conditional probabilities for sample 2000 / 2000
[t-SNE] Mean sigma: 1.709777
[t-SNE] KL divergence after 250 iterations with early exaggeration: 70.112206
[t-SNE] KL divergence after 1000 iterations: 1.132835
```

#TRAINING THE MODEL

```
(1600, 39)
(1600,)
(200, 39)
(200,)
(200, 39)
(200,)
(None, None)
```

#AFTER TRAINING THE DATA , COMPARING WITH THE RESPECTIVE K VALUE TO SEE THE ACCURACY OF THE MODEL

```
Accuracy with k=5 93.5
Accuracy with k=1 94.5
Accuracy with k=10 94.5
```

#CALCULATES AND PRINTS THE CONFUSION MATRICES FOR THREE DIFFERENT SETS OF PREDICTIONS (Y_PRED_1, Y_PRED_5, Y_PRED_10) COMPARED TO THE TRUE LABELS Y_VALID.

```
[[24  0  0  0  0  0  0  0]
 [ 0 44  1  0  0  0  0  0]
 [ 0  0 36  1  0  0  0  0]
 [ 1  1  0 12  5  0  0  0]
 [ 0  0  0  1 48  0  0  0]
 [ 0  0  0  0  0  2  0  0]
 [ 0  0  0  1  0  0 22  0]
 [ 0  0  0  0  0  0  0  1]]
[[24  0  0  0  0  0  0  0]
 [ 0 45  0  0  0  0  0  0]
 [ 0  0 37  0  0  0  0  0]
 [ 2  0  0  7 10  0  0  0]
 [ 0  0  0  0 49  0  0  0]
 [ 0  0  0  0  0  2  0  0]
 [ 0  0  0  0  0  0 23  0]
 [ 0  1  0  0  0  0  0  0]]
[[24  0  0  0  0  0  0  0]
 [ 0 44  1  0  0  0  0  0]
 [ 0  0 36  1  0  0  0  0]
 [ 1  1  0 12  5  0  0  0]
 [ 0  0  0  1 48  0  0  0]
 [ 0  0  0  0  0  2  0  0]
 [ 0  0  0  1  0  0 22  0]
 [ 0  0  0  0  0  0  0  1]]
```

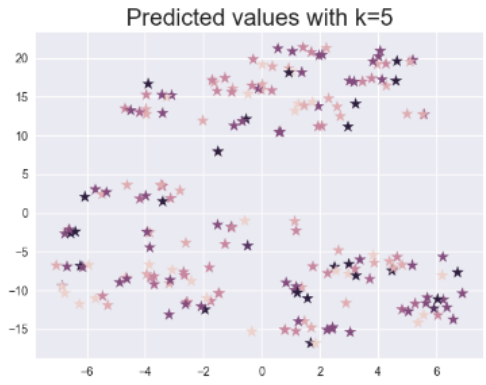
#CLASSIFICATION_REPORT

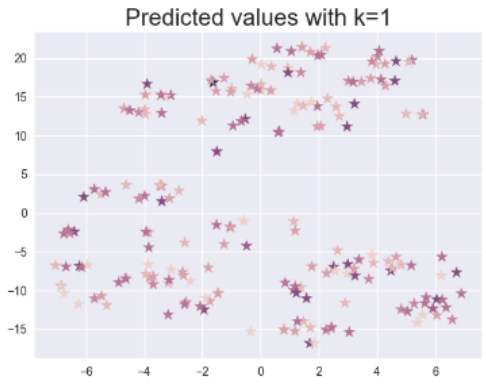
	precision	recall	f1-score	support
0	0.96	1.00	0.98	24
1	0.98	0.98	0.98	45
2	0.97	0.97	0.97	37
3	0.80	0.63	0.71	19
4	0.91	0.98	0.94	49
5	1.00	1.00	1.00	2
6	1.00	0.96	0.98	23
9	1.00	1.00	1.00	1
accuracy			0.94	200
macro avg	0.95	0.94	0.94	200
weighted avg	0.94	0.94	0.94	200

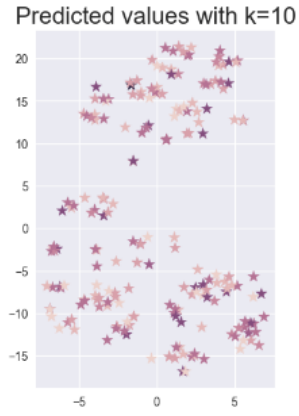
	precision	recall	f1-score	support
0	0.92	1.00	0.96	24
1	0.98	1.00	0.99	45
2	1.00	1.00	1.00	37
3	1.00	0.37	0.54	19
4	0.83	1.00	0.91	49
5	1.00	1.00	1.00	2
6	1.00	1.00	1.00	23
9	0.00	0.00	0.00	1
accuracy			0.94	200
macro avg	0.84	0.80	0.80	200
weighted avg	0.94	0.94	0.92	200

	precision	recall	f1-score	support
0	0.96	1.00	0.98	24
1	0.98	0.98	0.98	45
2	0.97	0.97	0.97	37
3	0.80	0.63	0.71	19
4	0.91	0.98	0.94	49
5	1.00	1.00	1.00	2
6	1.00	0.96	0.98	23
9	1.00	1.00	1.00	1
accuracy			0.94	200
macro avg	0.95	0.94	0.94	200
weighted avg	0.94	0.94	0.94	200

#SCATTER PLOT FOR THE PREDICTED VALUES







#MERGING THE X_TEST VALUE TO THE ACTUAL DATASET

[98]:

	artist	song	duration_ms	explicit	year	popularity	danceability	energy	key	loudness	...	Folk/Acoustic	latin	rock	jazz	country	country	easy listening	blues	cat	cluster
0	Sisqo	Thong Song	0.379185	1.0	1999	0.775281	0.682033	0.882428	0.181818	0.669780	...	0	0	0	0	0	0	0	0	0.333333	1
1	Eminem	The Real Slim Shady	0.461274	1.0	2000	0.966292	0.969267	0.641987	0.454545	0.803933	...	0	0	0	0	0	0	0	0	0.222222	1
2	Modjo	Lady - Hear Me Tonight	0.523118	0.0	2001	0.865169	0.698582	0.797691	0.545455	0.735596	...	0	0	0	0	0	0	0	0	0.222222	5
3	Melanie C	Never Be The Same Again	0.488218	0.0	1999	0.685393	0.661939	0.667408	0.272727	0.759018	...	0	0	0	0	0	0	0	0	0.222222	0
4	Alice Deejay	Better Off Alone	0.274509	0.0	2000	0.820225	0.640662	0.873954	0.727273	0.709803	...	0	0	0	0	0	0	0	0	0.666667	4
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
209	Drake	Money In The Grave (Drake ft. Rick Ross)	0.249029	1.0	2019	0.853933	0.829787	0.473573	0.909091	0.813766	...	0	0	0	0	0	0	0	0	0.111111	1
210	Anuel AA	China	0.508463	0.0	2019	0.853933	0.776596	0.797691	0.636364	0.830714	...	0	0	0	0	0	0	0	0	0.222222	0
211	Flipp Dinero	Leave Me Alone	0.222654	1.0	2019	0.775281	0.783688	0.728842	0.636364	0.874988	...	0	0	0	0	0	0	0	0	0.111111	1
212	Ed Sheeran	Take Me Back to London (feat. Stormzy)	0.206746	1.0	2019	0.741573	0.893617	0.748967	0.727273	0.741229	...	0	0	0	0	0	0	0	0	0.444444	4
213	Sam Feldt	Post Malone (feat. RANI)	0.165552	0.0	2019	0.775281	0.544917	0.621862	0.636364	0.822413	...	0	0	0	0	0	0	0	0	0.444444	0

214 rows × 45 columns

#FINAL_OUTPUT

```
Enter the song name:  Money In The Grave (Drake ft. Rick Ross)
Enter the artist name:  Drake
Song name:  Money In The Grave (Drake ft. Rick Ross)
Artist name:  Drake
Song Selected:- Money In The Grave (Drake ft. Rick Ross) Index:  209
Searching for recommendations.....
1042                                    Hard      Rihanna
1107    Who's That Chick? (feat. Rihanna)      David G...
944                                    Sober      P!nk
1199                A Thousand Years      Christina Perri
331     Damn! (feat. Lil' Jon) - Club Mix      Youngbl...
dtype: object
```