

What makes a hit song?

Fadi Sarraf

Mahima Kaushiva

Tina Liu

Target Customer

- The blueship record companies
- The independent music publishers

That employ talent scouts in an aim to spot the next Billboard hit song, support existing artists in making new songs.

Knowing current **music trends** are key to the success of these companies.



SONY MUSIC



WARNER MUSIC GROUP

Data

- Source: Spotify API
- Consists of over 200k randomly selected tracks released between 1999 and 2019

For each song, we have info on:

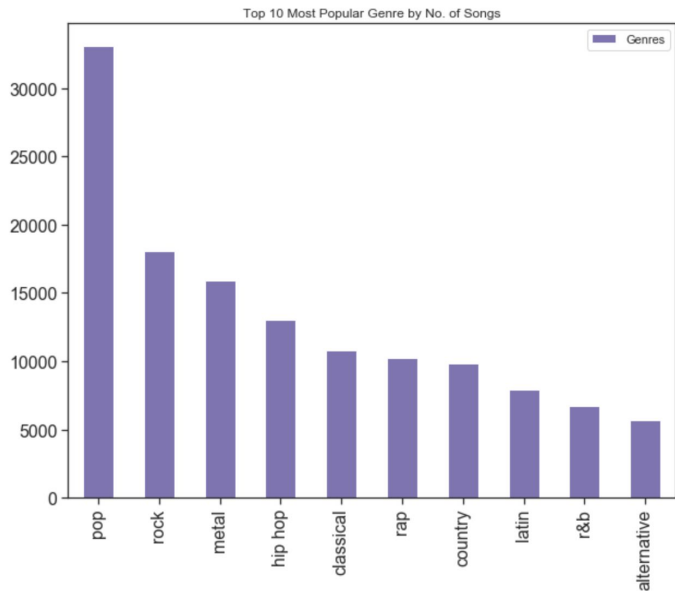
- Popularity score (0 to 100)
- Artist popularity
- Genre and Subgenre
- Composition info like:
 - Acousticness
 - Danceability
 - Energyas defined by Spotify

In our study we will aim to

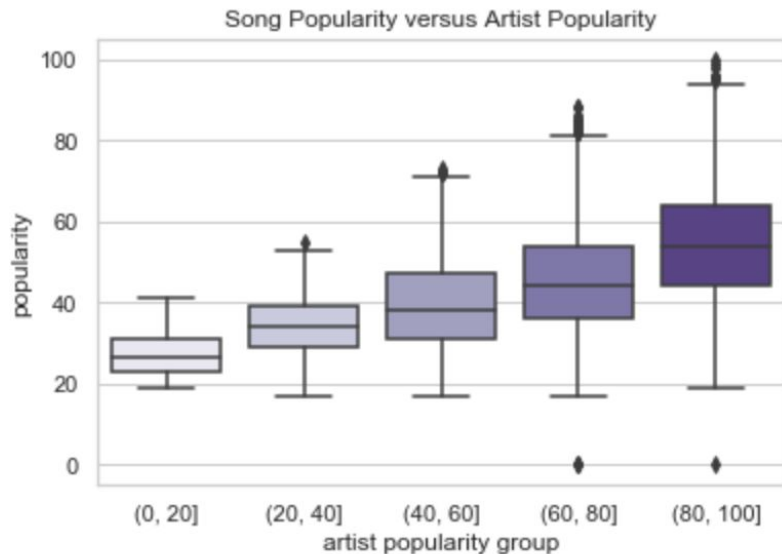
predict the popularity of song
and reveal other useful insights.

The Most Popular...

- Twice as many pop songs were made as the next “popular” genre being rock
- Most popular Genre is edm and pop
- Least popular Genre is jazz and folk



- being a popular artist **does not guarantee** a hit song



Predictive Model

Conclusion

we cannot use characteristic about the song to predict the popularity of the song accurately

but having
- a popular artist
- being louder
- being more acoustic
- being shorter in duration
makes a **Rap** song more popular

In building a predictive (linear) model, we have tried to:

1) Segment songs by genre

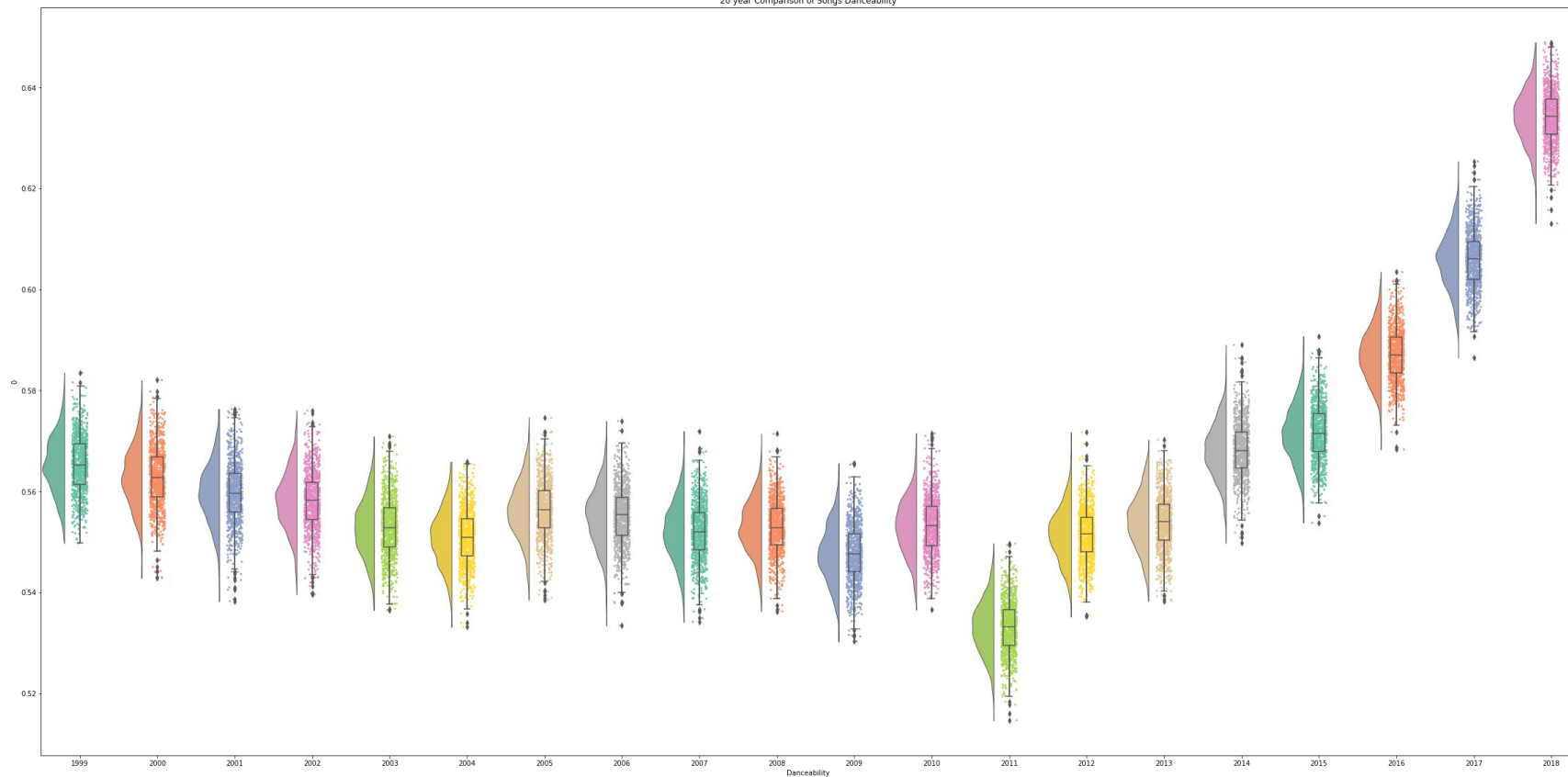
- a) For **R&B and Rap**, the R^2 is relatively higher at around **0.4 level**
- b) For **pop and rock**, R^2 only around **0.20**
- c) Polynomial transformation improves R^2 to 0.45 for Rap, no material drop noted for R^2 when run on test dataset for Rap.

2) Segment songs by sub-genre & by year and genre

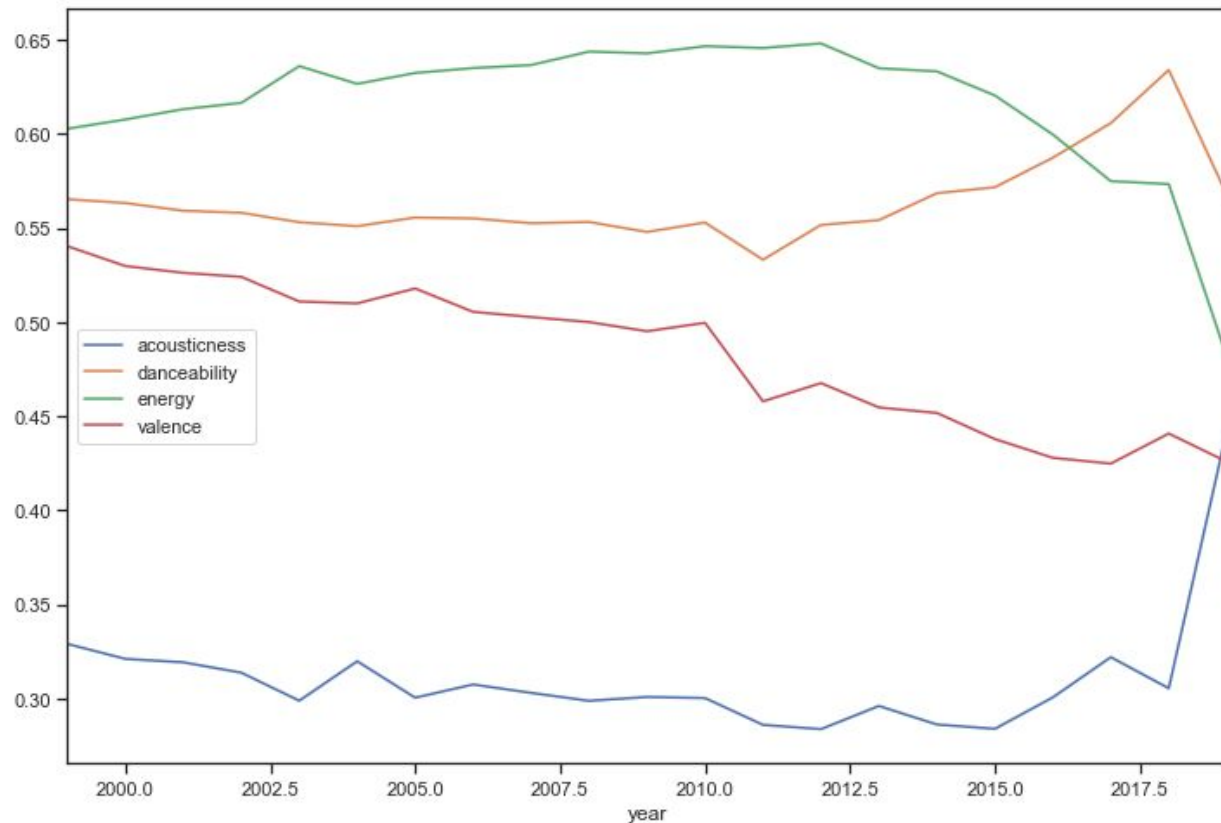
- a) Further segment by year gives even less predictive power
- b) We get similar low R^2 on most sub-segments, whenever we get a higher R^2 (**0.6 region**) the data sample is small ~ 200 . For example, "electro house".

Trends on Song Danceability

20 year Comparison of Songs Danceability



- Post 2010, songs have been getting less energetic
- Since last year, danceability of song has seen a sharp decrease while songs are getting much more acoustic
- Over time, songs are getting more “dark” slowly



Post 2000 Trends

Recommendations

Unfortunately, for our target customer

- We cannot exactly breakdown the ingredients to a hit song as popularity of a song is **not strongly dependent** on characteristic of a song.
- However, noting on recent trends...
 - ◆ on average songs are getting more dark, less danceable and more acoustic
- Also, a **popular artist can still turn out unpopular songs**

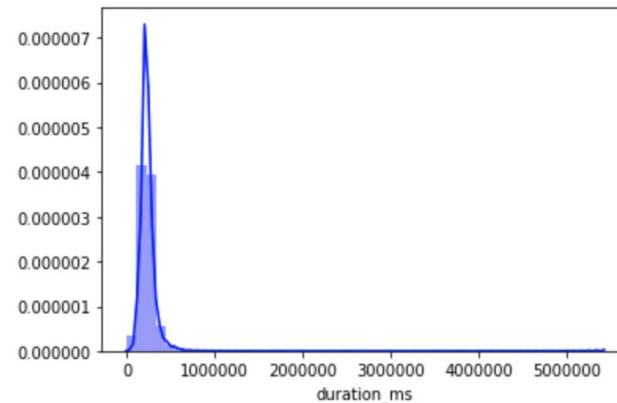
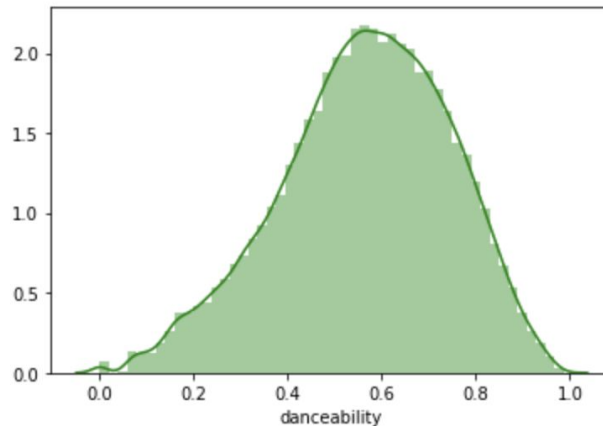
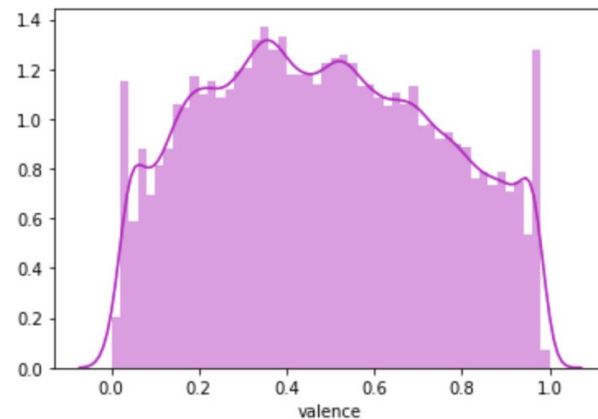
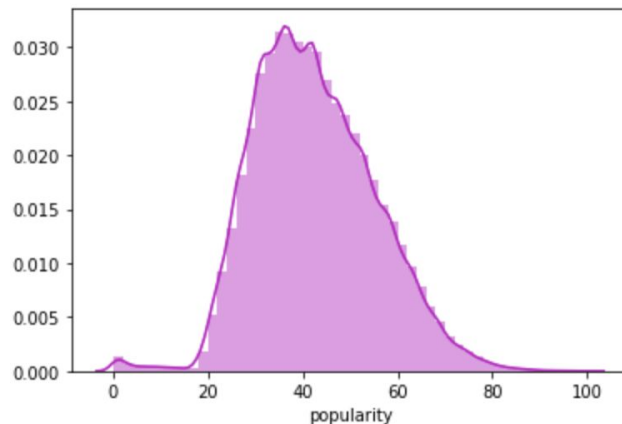
Further Study

We could explore using market effect measures like number of mention/hashtags on social media to try to predict the popularity of a song.

Appendix

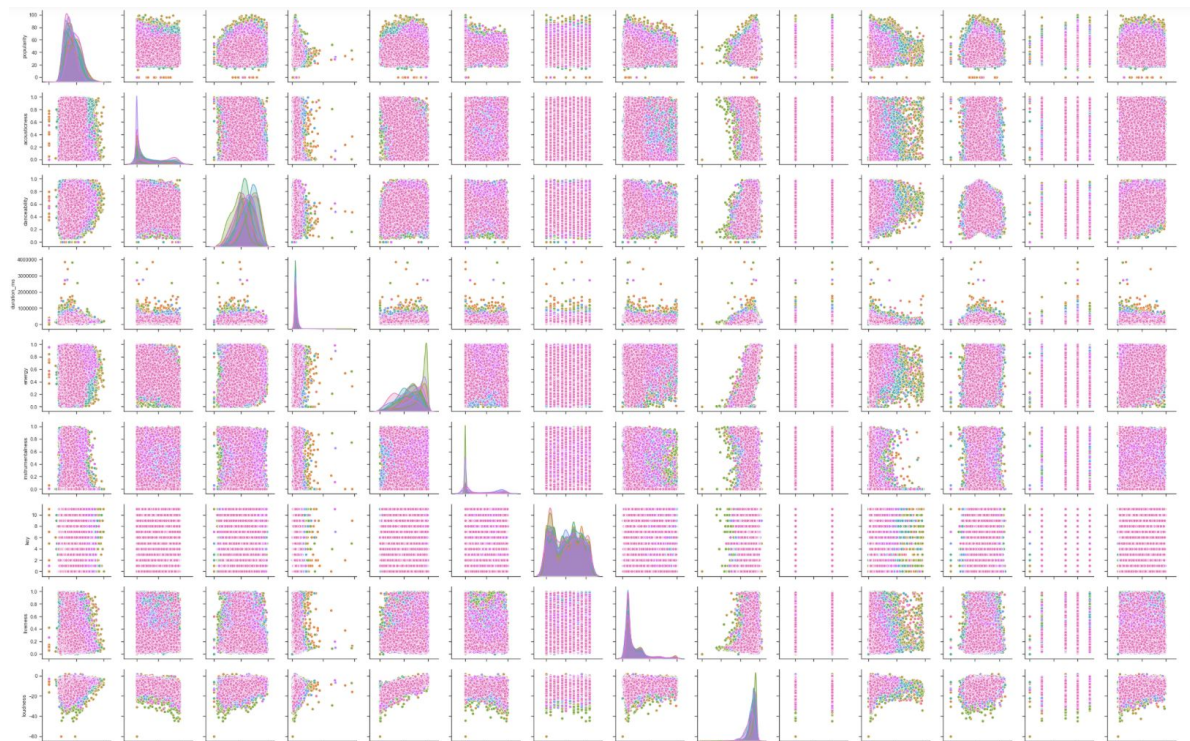
Distribution Check

Note that the characteristic info about a song is on scale of 0 to 1 so these seem to have been scaled by Spotify already.



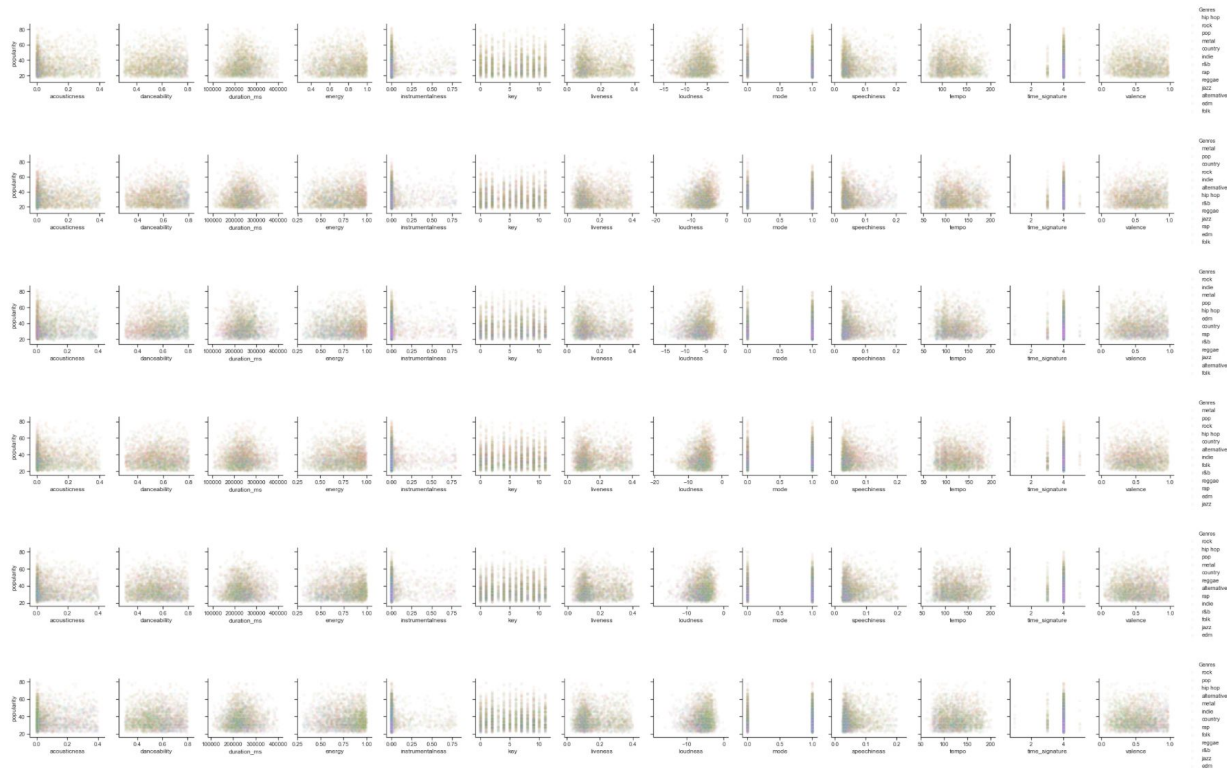
→ We take songs between 1.7 minutes to 6.7 minutes

Pairplots: Just noise!

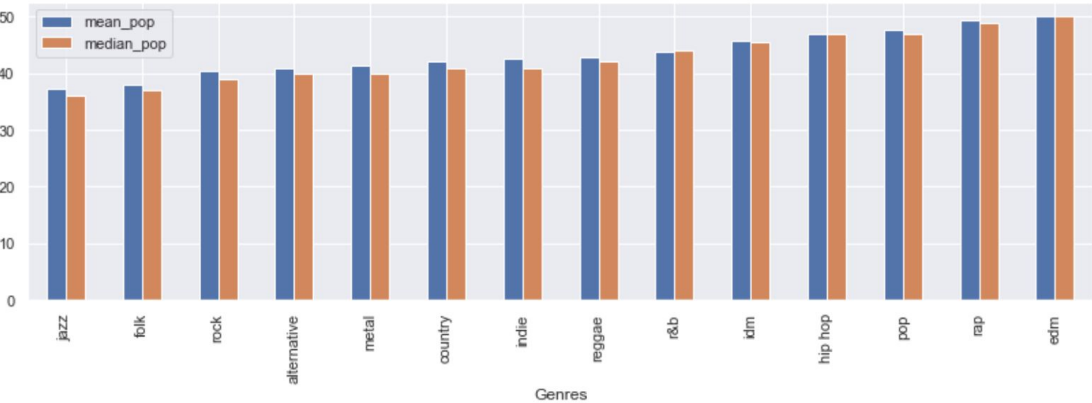


Pairplots: (including year and genre)

more noise!



Popularity by Genre



		mean_pop	median_pop
Genres			
jazz	37.327035	36.0	
folk	38.055092	37.0	
rock	40.395254	39.0	
alternative	40.883949	40.0	
metal	41.466558	40.0	
country	42.162795	41.0	
indie	42.708052	41.0	
reggae	42.950216	42.0	
r&b	43.910603	44.0	
idm	45.816092	45.5	
hip hop	46.987184	47.0	
pop	47.601078	47.0	
rap	49.412992	49.0	
edm	50.069478	50.0	

Model Results: Data split by Genre

Coef for Rap songs linear model ->

```
Index(['acousticness', 'danceability', 'duration_ms', 'energy',  
      'instrumentalness', 'liveness', 'loudness', 'mode', 'speechiness',  
      'tempo', 'valence', 'popularity_artist'],  
      dtype='object')  
array([ 1.25183529e+01, -1.15453462e+01, -6.13375479e-05, -6.51802777e+00,  
       -7.57379141e-01, -6.18874397e+00,  2.89919027e-01, -7.78839965e-01,  
       -1.60091597e+01,  2.01908207e-02, -6.16019617e+00,  3.97667392e-01])
```

	genres	lin_train_mse	lin_test_mse	lin_r2	lin_test_r2	poly_mse	test_mse	poly_r2	test_poly_r2
7	rap	108.274259	108.166315	0.377743	0.388960	94.014777	108.166315	0.459693	0.401466
11	edm	93.679935	100.833025	0.334453	0.329323	82.449667	100.833025	0.414238	0.344331
0	hip hop	130.335753	130.247358	0.273583	0.265256	119.912186	130.247358	0.331678	0.257997
6	r&b	107.689442	130.050462	0.380412	0.340285	104.963996	130.050462	0.396093	0.250596
2	pop	135.727368	131.024886	0.224507	0.218118	131.919561	131.024886	0.246264	0.230452
3	metal	93.665194	87.402356	0.194839	0.199596	89.078852	87.402356	0.234264	0.230286
4	country	131.487849	134.052844	0.182822	0.183073	120.724259	134.052844	0.249716	0.213753
1	rock	113.634359	106.974171	0.158831	0.141324	106.163680	106.974171	0.214132	0.163737
10	alternative	83.181637	83.648094	0.182100	0.192306	77.243530	83.648094	0.240488	0.154775
5	indie	97.440913	118.867878	0.186428	0.153221	101.236556	118.867878	0.154737	-0.071593
9	jazz	80.225034	117.637524	0.104469	0.100275	57.939424	117.637524	0.353238	-0.099938
8	reggae	105.629688	129.872114	0.138631	0.100718	79.237189	129.872114	0.353851	-0.157534
12	folk	82.400121	127.976106	0.032738	-0.036609	64.199234	127.976106	0.246391	-0.414730

Model Results:

Data split by Genre & Year

<- overfitting

	year	genres	lin_train_mse	lin_test_mse	lin_r2	lin_test_r2	poly_mse	test_mse	poly_r2	test_poly_r2
142	2018	pop	52.903743	57.019409	0.253634	0.218482	43.804276	57.019409	0.382009	0.117779
11	2001	metal	115.453343	192.161653	0.218271	0.211946	89.486520	192.161653	0.394091	0.002493
39	2006	rock	91.711362	119.282539	0.183191	0.169330	66.335307	119.282539	0.409198	-0.015842
3	1999	metal	94.695424	167.266054	0.214866	0.277767	63.301129	167.266054	0.475161	-0.016667
102	2014	pop	74.422688	82.509144	0.134197	0.122725	65.210377	82.509144	0.241369	-0.048605
92	2013	pop	77.622257	92.570019	0.172674	0.176136	64.316884	92.570019	0.314487	-0.077952
84	2012	pop	88.047274	118.426069	0.186157	0.127120	72.246890	118.426069	0.332204	-0.122128
132	2017	pop	51.450971	80.773169	0.187972	0.130980	48.253839	80.773169	0.238431	-0.165097
35	2005	metal	88.990197	152.246767	0.231579	0.137005	67.701926	152.246767	0.415401	-0.171773
122	2016	pop	55.482429	83.387603	0.190843	0.135759	45.847260	83.387603	0.331363	-0.182692