

# Spotify Music Popularity



Qingyin Ge  
qg2177

Data  
Overview

Data  
Description

Data  
Analysis

# Data Description

Total Observations: 2476

Variables:

- 1 target variable: Popularity
- 9 Numeric variables: BPM, Energy, Danceability, Loudness, Liveness, Valence, Duration, Acousticness, Speechiness
- Billboard variables: Title, Artist, Year, Genre

Sample  
Data



	Title	Artist	Top.Genre	Year	BPM	Energy	Danceability	Loudness	Liveness	Valence	Duration	Acousticness	Speechiness	Popularity
1951	Blue Suede C...	Elvis Presley	adult standar...	1956	95	66	56	-8	14	96	7	65	6	62
1952	Jailhouse Rock	Elvis Presley	adult standar...	1958	167	58	65	-10	7	92	26	41	8	73
1953	My Baby Just ...	Nina Simone	adult standar...	1958	119	24	75	-15	9	43	96	83	12	68
1954	Heartbreak H...	Elvis Presley	adult standar...	1958	94	21	70	-12	11	72	14	84	7	63
1955	Hound Dog	Elvis Presley	adult standar...	1958	175	76	36	-8	76	95	18	73	6	69
1956	Johnny B. Go...	Chuck Berry	blues rock	1959	168	80	53	-9	31	97	42	74	7	74
1957	Take Five	The Dave Bru...	bebop	1959	174	26	45	-13	7	60	203	54	4	65
1958	Blueberry Hill	Fats Domino	adult standar...	1959	133	50	49	-10	16	83	28	74	3	56
1798	At Last	Etta James	adult standar...	1960	87	35	27	-9	33	33	60	55	3	74
1799	Can't Help Fa...	Elvis Presley	adult standar...	1961	100	29	40	-14	11	34	62	94	3	78
1800	Fly Me To Th...	Frank Sinatra	adult standar...	1962	119	26	67	-14	6	37	27	45	5	77
1801	Stand by Me	Ben E. King	adult standar...	1962	118	31	65	-9	7	61	60	57	4	77
1802	Oh, Pretty Wo...	Roy Orbison	adult standar...	1962	127	60	62	-9	7	96	59	72	3	73
1803	Ring of Fire	Johnny Cash	arkansas cou...	1963	104	59	66	-8	35	78	38	62	3	71
1804	Blowin' in the...	Bob Dylan	album rock	1963	175	10	38	-21	6	44	45	91	5	69
1805	In Dreams	Roy Orbison	adult standar...	1963	108	44	50	-10	13	71	49	67	3	57
1806	Twist And Sh...	The Beatles	british invasion	1963	125	85	48	-9	4	94	35	64	5	74
1807	Are You Lone...	Elvis Presley	adult standar...	1963	66	9	36	-18	11	19	67	92	5	59
1808	I Saw Her Sta...	The Beatles	british invasion	1963	160	80	49	-10	7	97	54	27	4	71
1809	Don't Think T...	Bob Dylan	album rock	1963	111	24	60	-17	25	47	97	89	4	65
1810	The Sound of...	Simon & Garf...	classic rock	1964	107	22	53	-14	11	33	65	84	3	75
1811	House Of The...	The Animals	album rock	1964	117	58	30	-7	10	23	151	0	3	69
1812	Folsom Priso...	Johnny Cash	arkansas cou...	1964	111	53	69	-10	28	78	36	19	3	58
1813	The Times Th...	Bob Dylan	album rock	1964	172	40	39	-8	8	59	72	89	3	69
1814	A Change Is ...	Sam Cooke	adult standar...	1964	174	38	21	-10	30	45	71	72	3	66
1815	A Hard Day's ...	The Beatles	british invasion	1964	139	81	59	-6	10	80	34	14	4	69
1816	Nathalie	Gilbert Bécaud	chanson	1964	135	63	42	-7	59	81	125	75	8	53
1817	Eight Days A ...	The Beatles	british invasion	1964	138	58	65	-8	12	74	44	41	4	66
1818	It's All Over N...	The Rolling S...	album rock	1964	99	69	60	-11	33	76	86	2	3	46
1819	Yesterday – R...	The Beatles	british invasion	1965	97	18	33	-12	9	32	12	88	3	78

# Data Description

Total Observations: 2476

Variables:

- 1 target variable: Popularity
- 9 Numeric variables: BPM, Energy, Danceability, Loudness, Liveness, Valence, Duration, Acousticness, Speechiness
- Billboard variables: Title, Artist, Year, Genre

Sample  
Data

# Spotify Music Popularity



Qingyin Ge  
qg2177

Data  
Overview

Data  
Description

Data  
Analysis

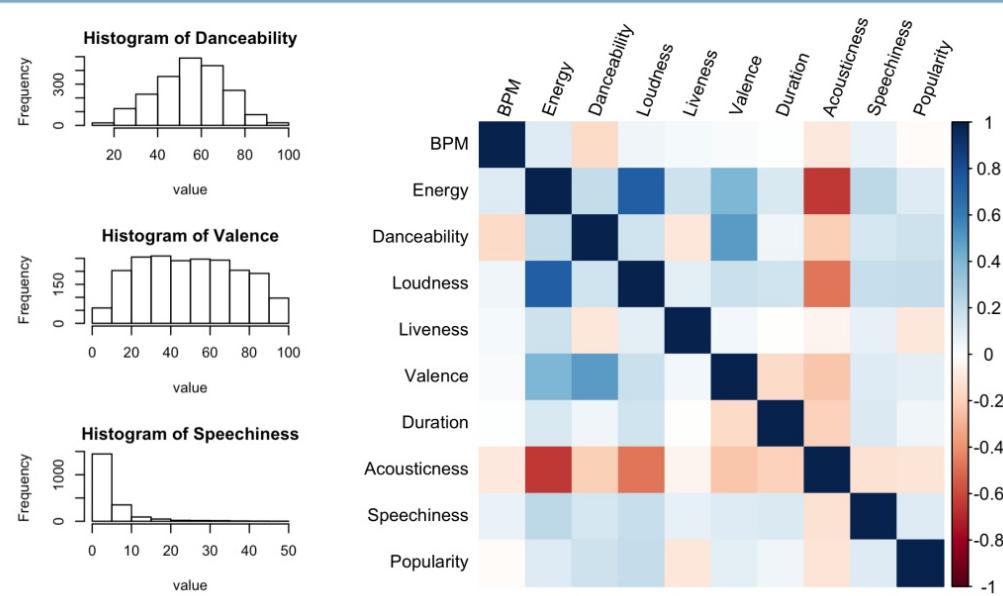
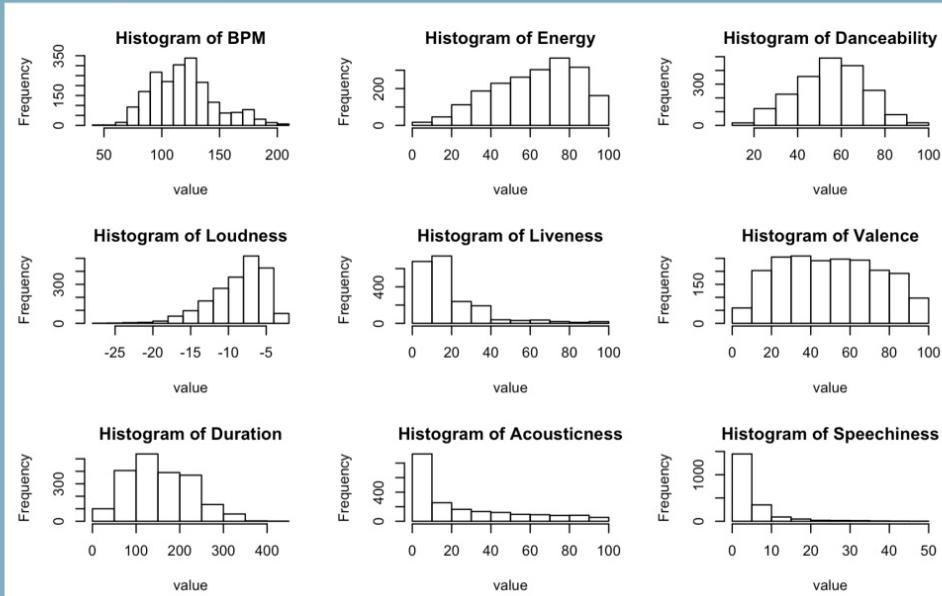
# Data Overview

If you don't  
know Statistics  
...

Numeric Features

Categorical Variables

# Distribution & Correlation



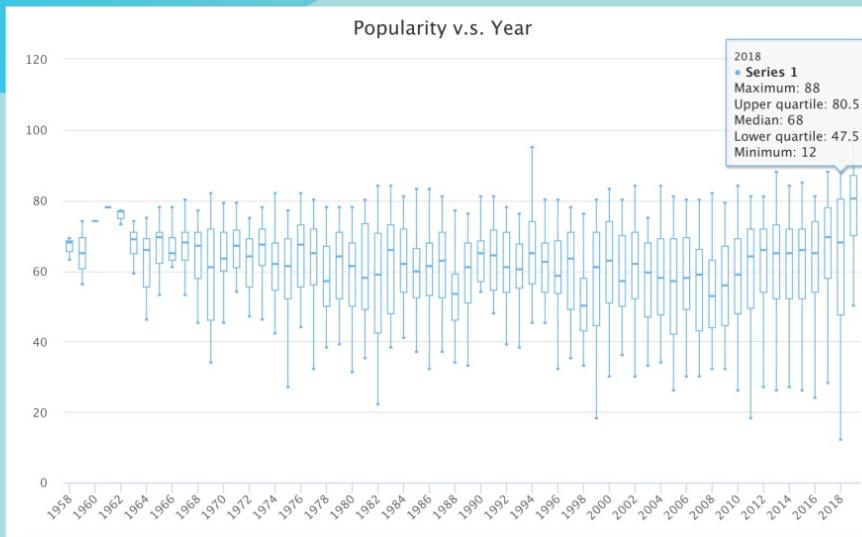
# Data Overview

If you don't  
know Statistics  
...

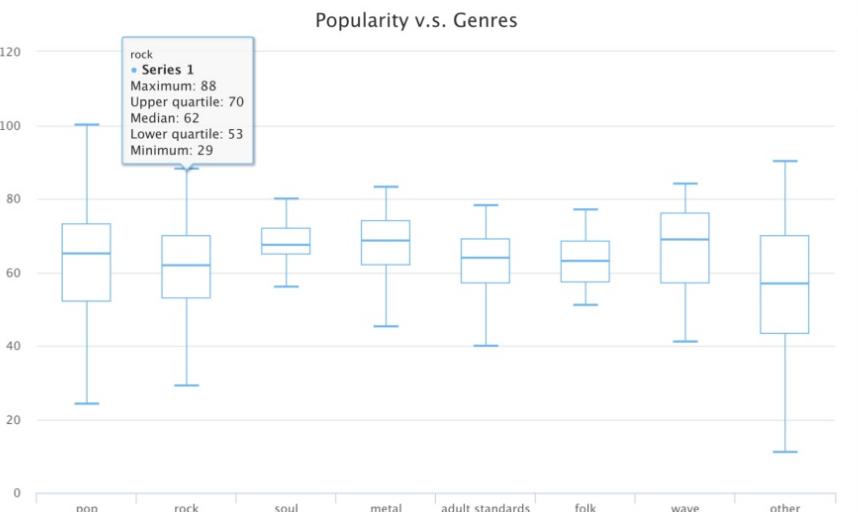
Numeric Features

Categorical Variables

# Popularity Change With...



# 70s

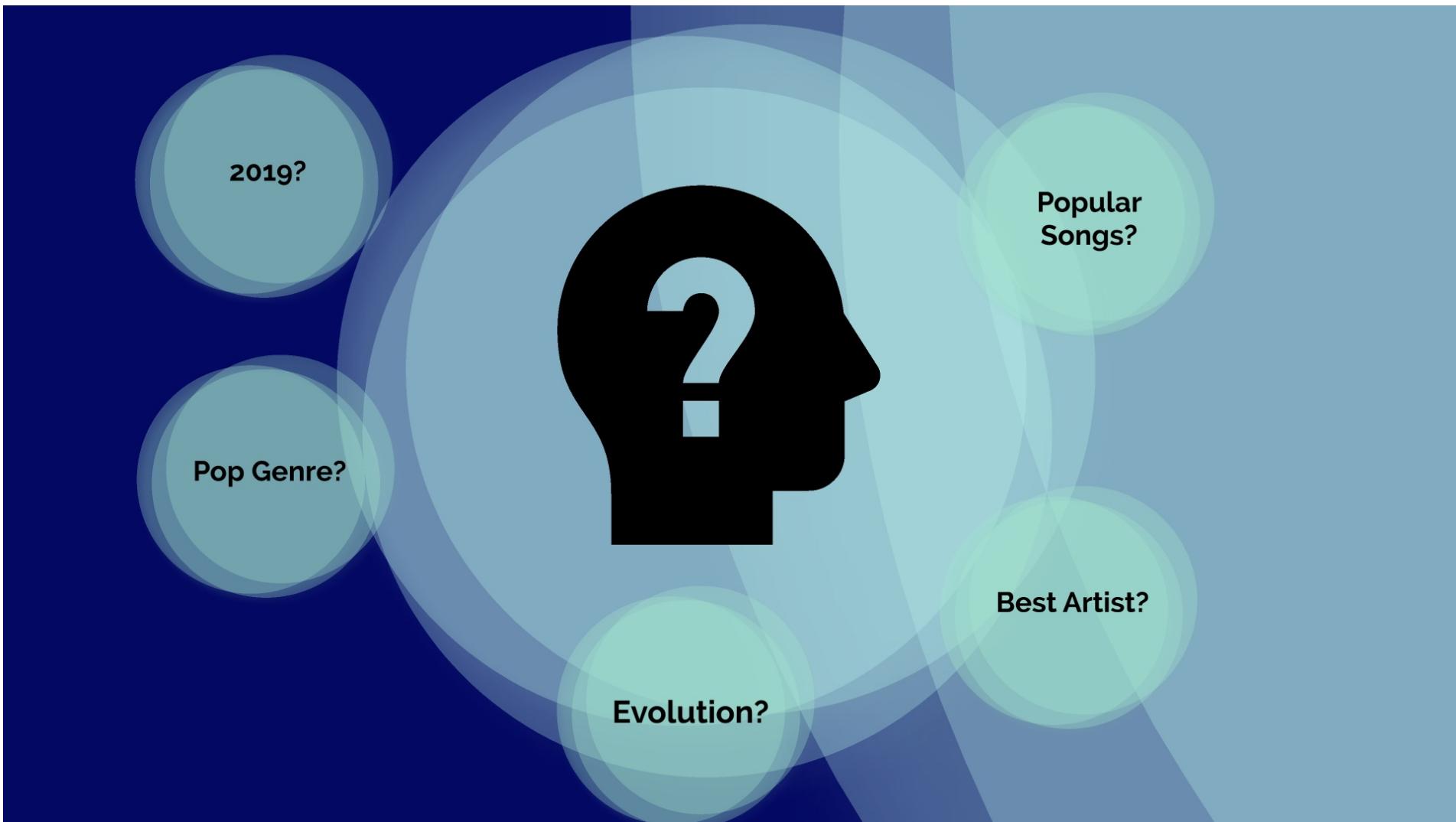


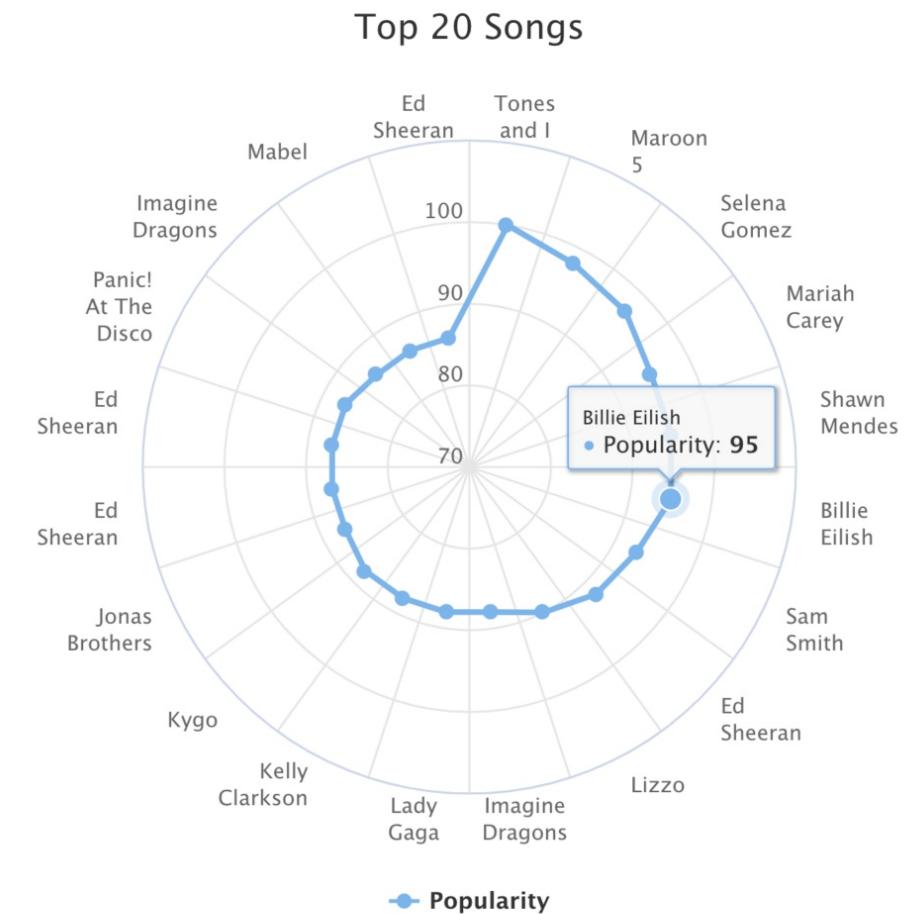
# Data Overview

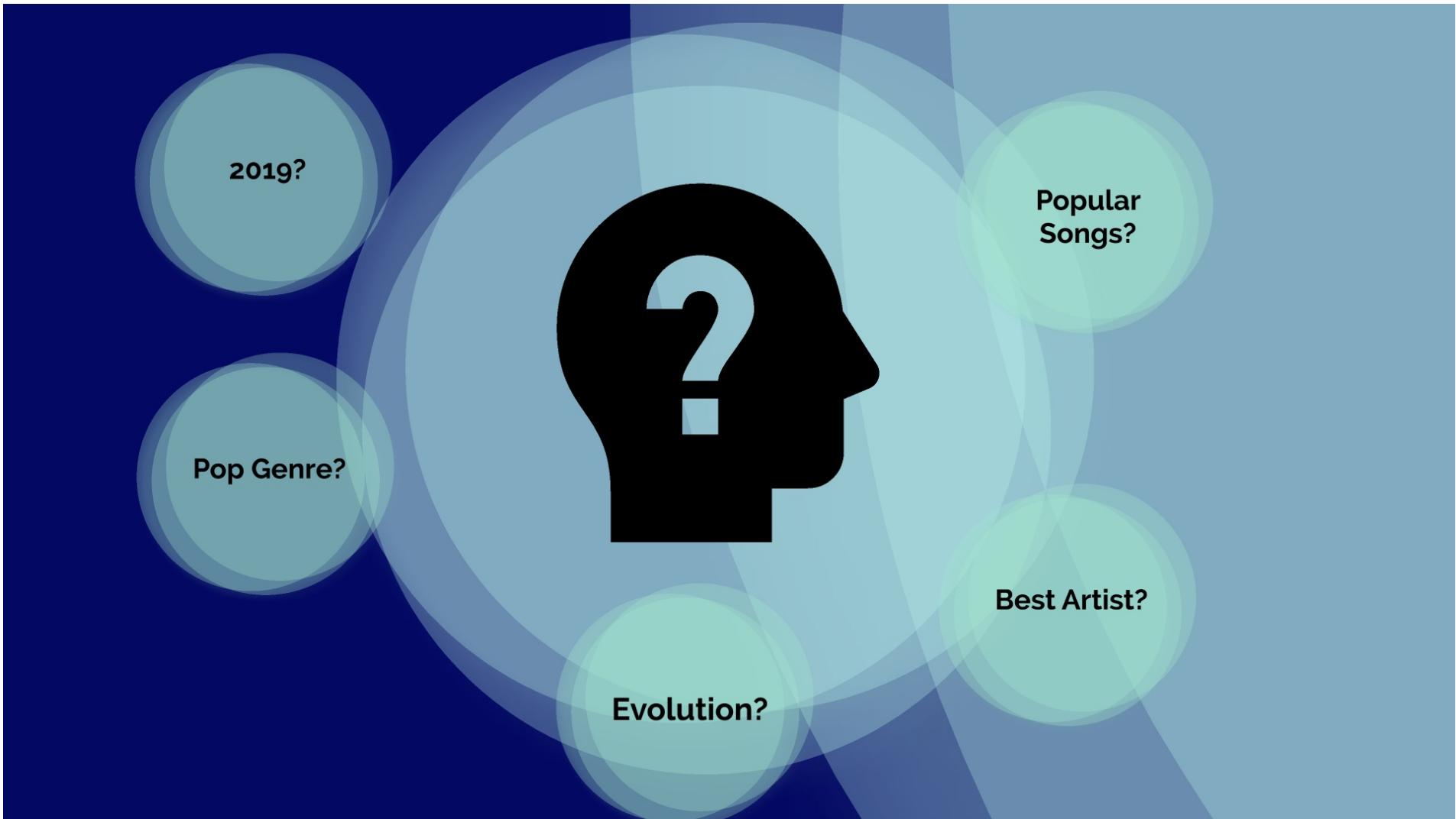
If you don't  
know Statistics  
...

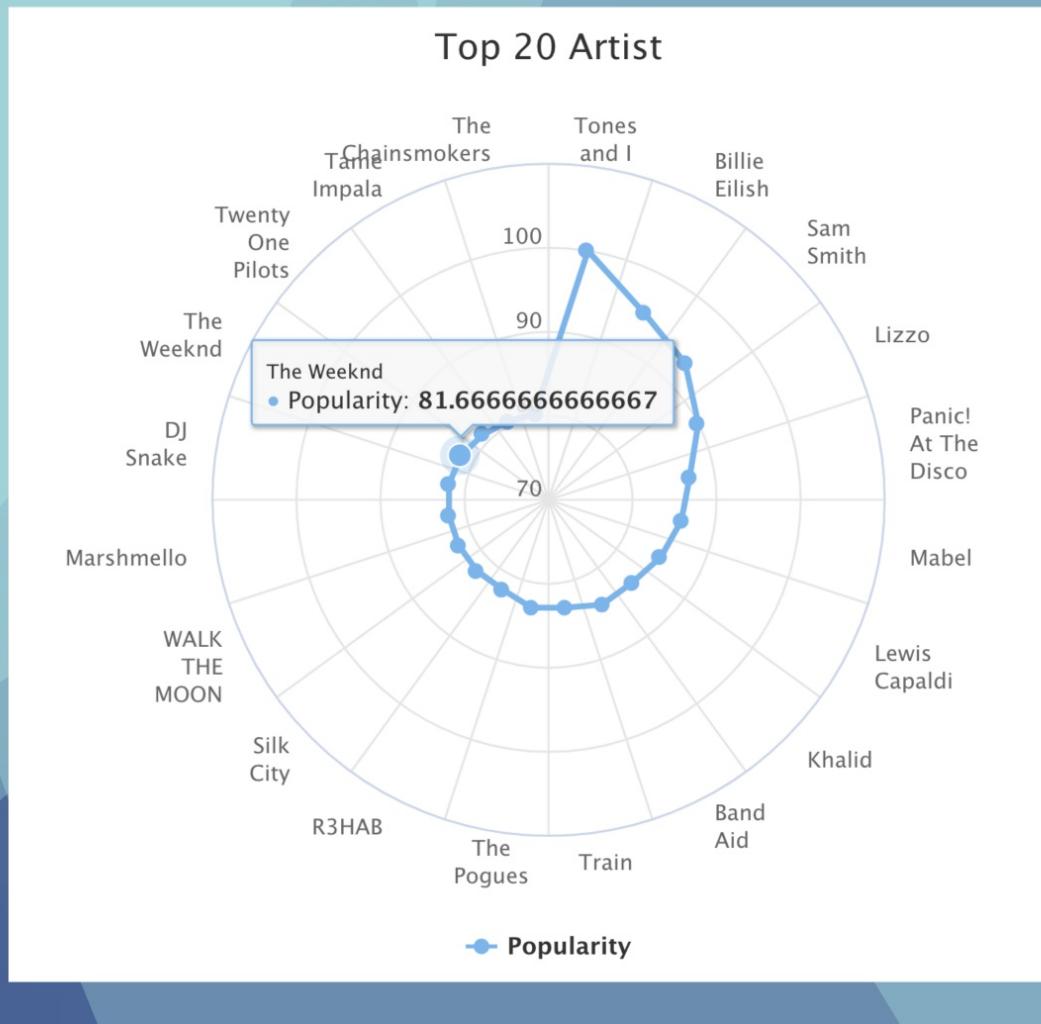
Numeric Features

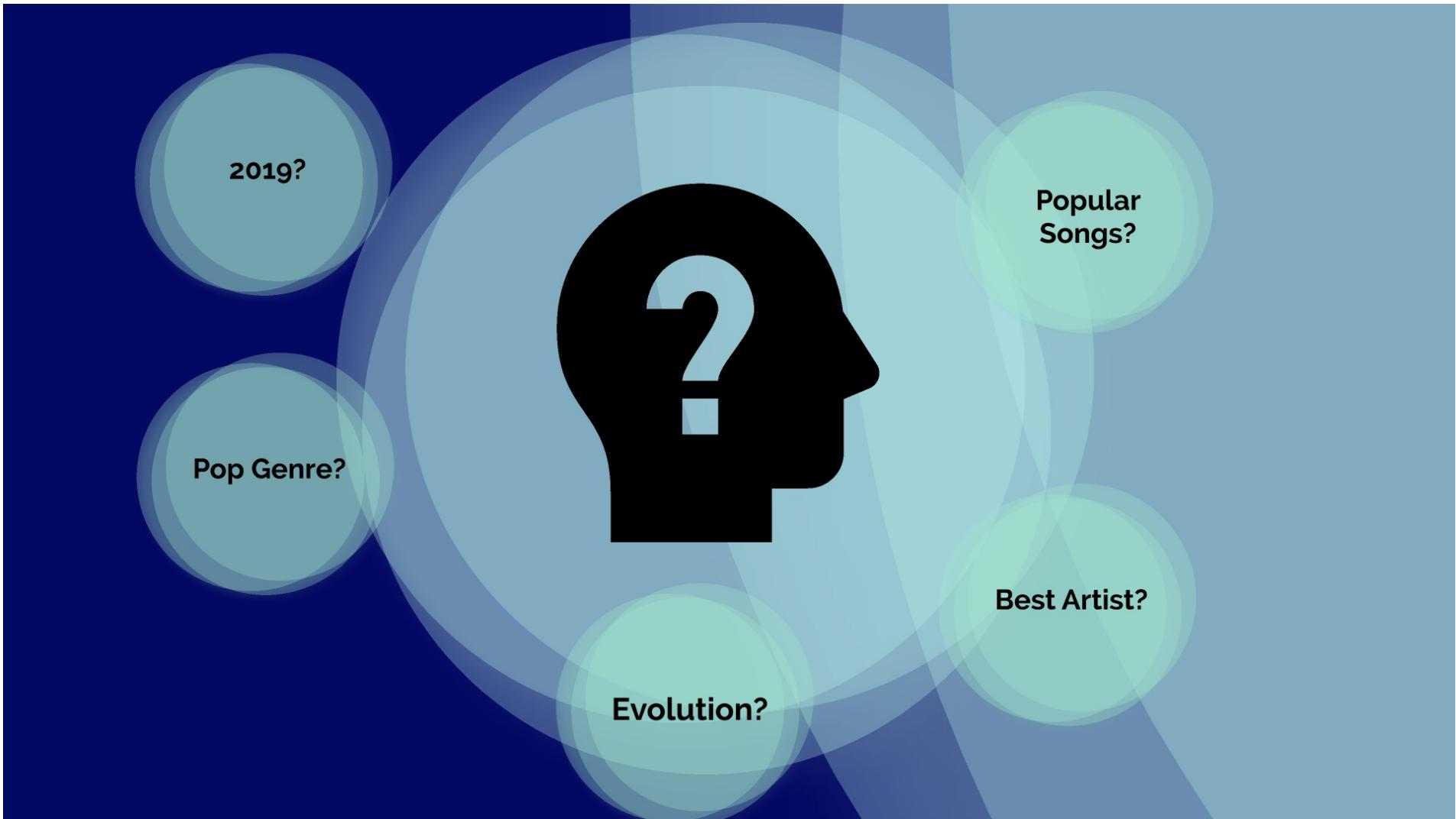
Categorical Variables

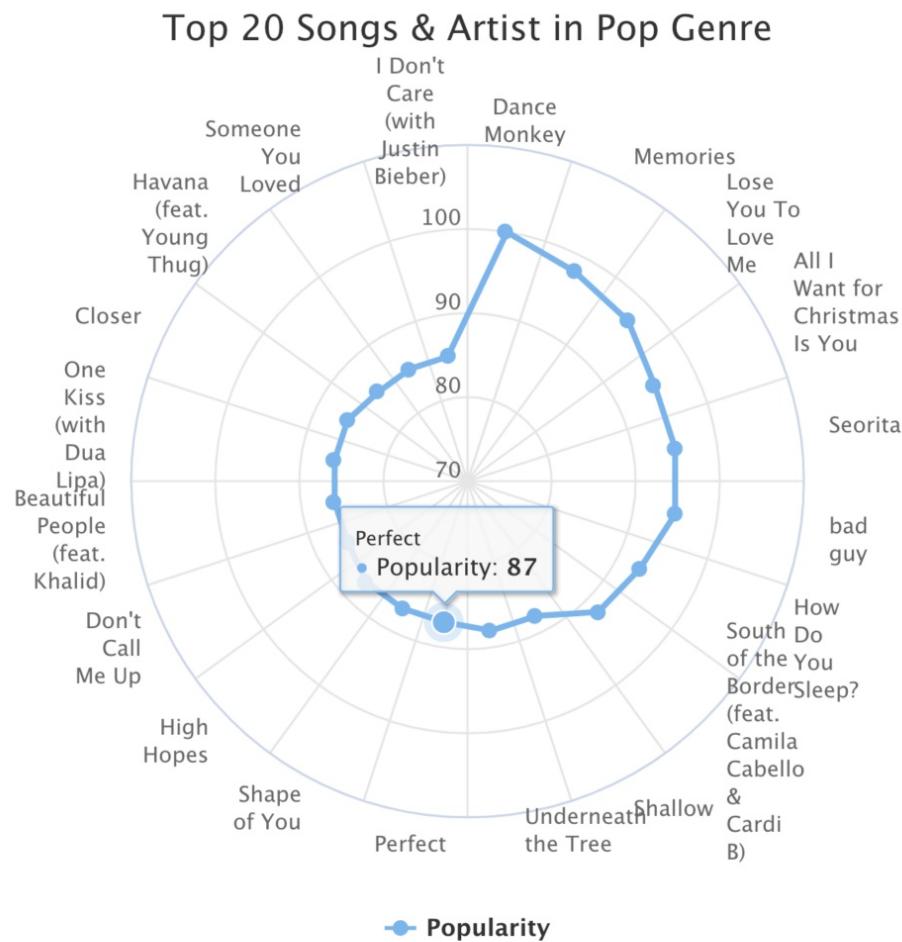


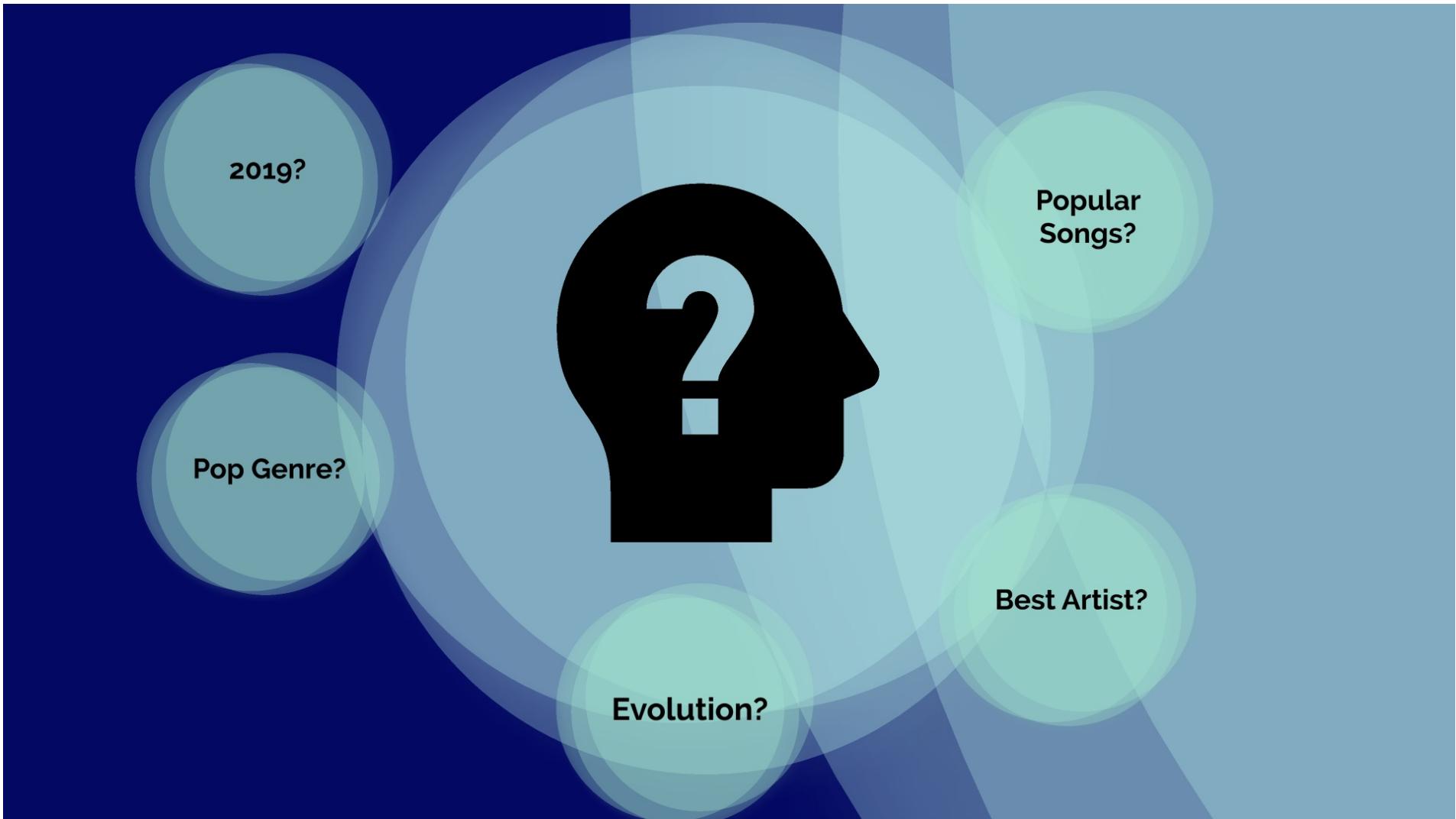


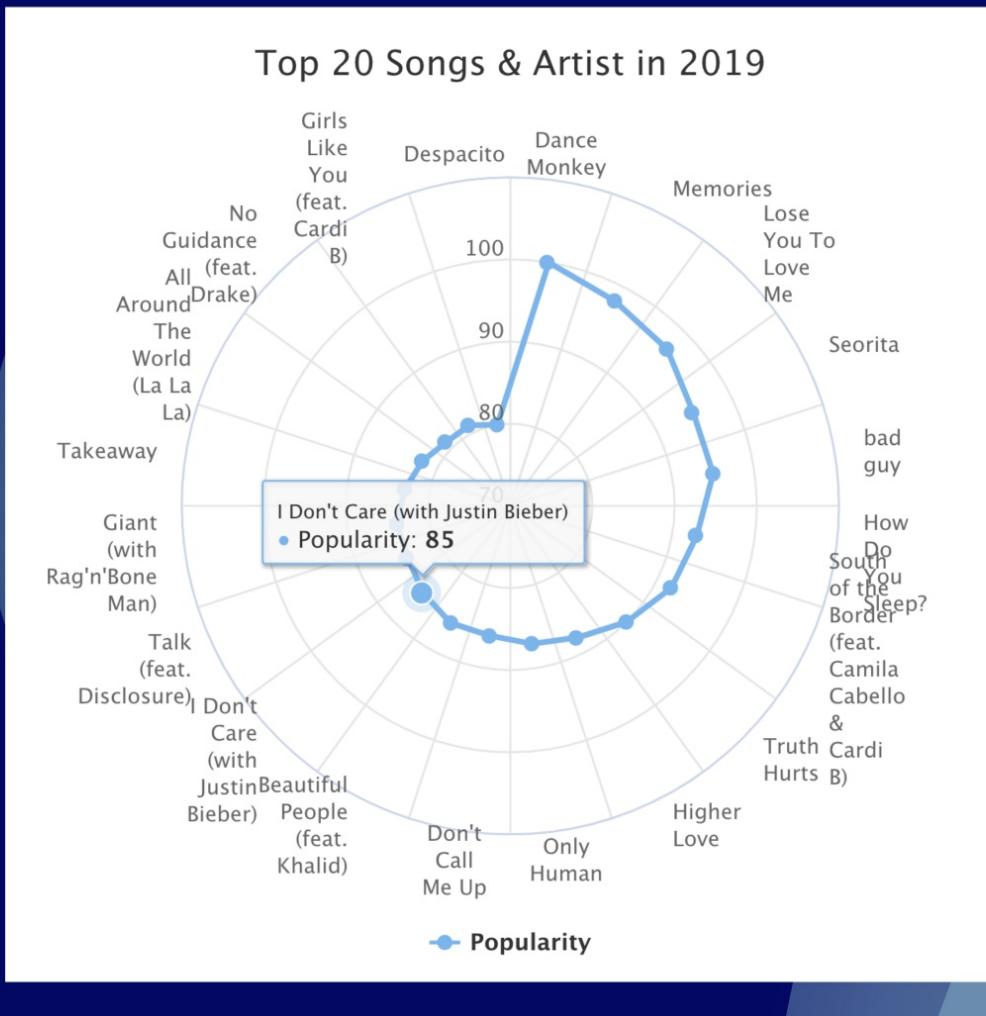


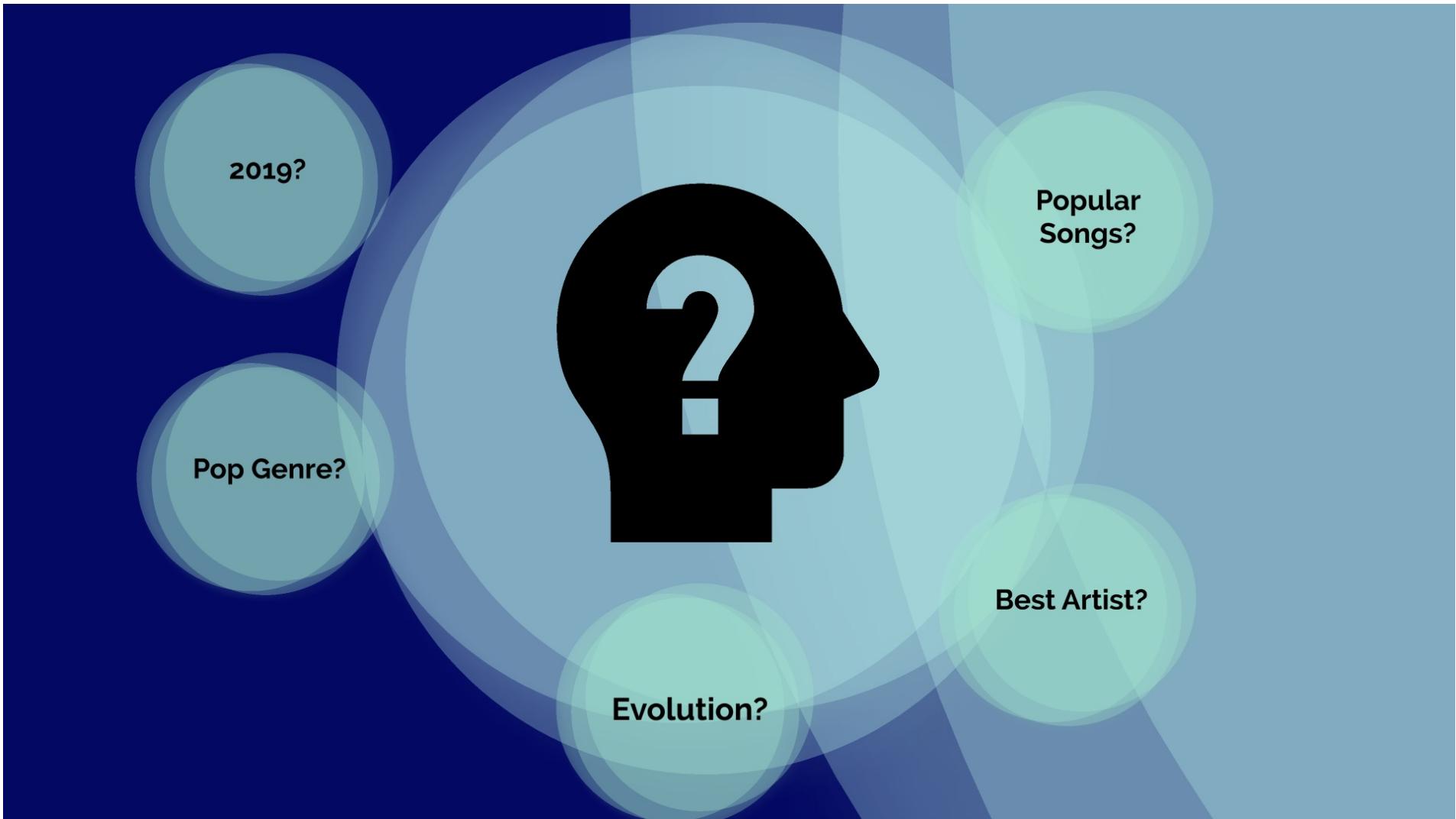




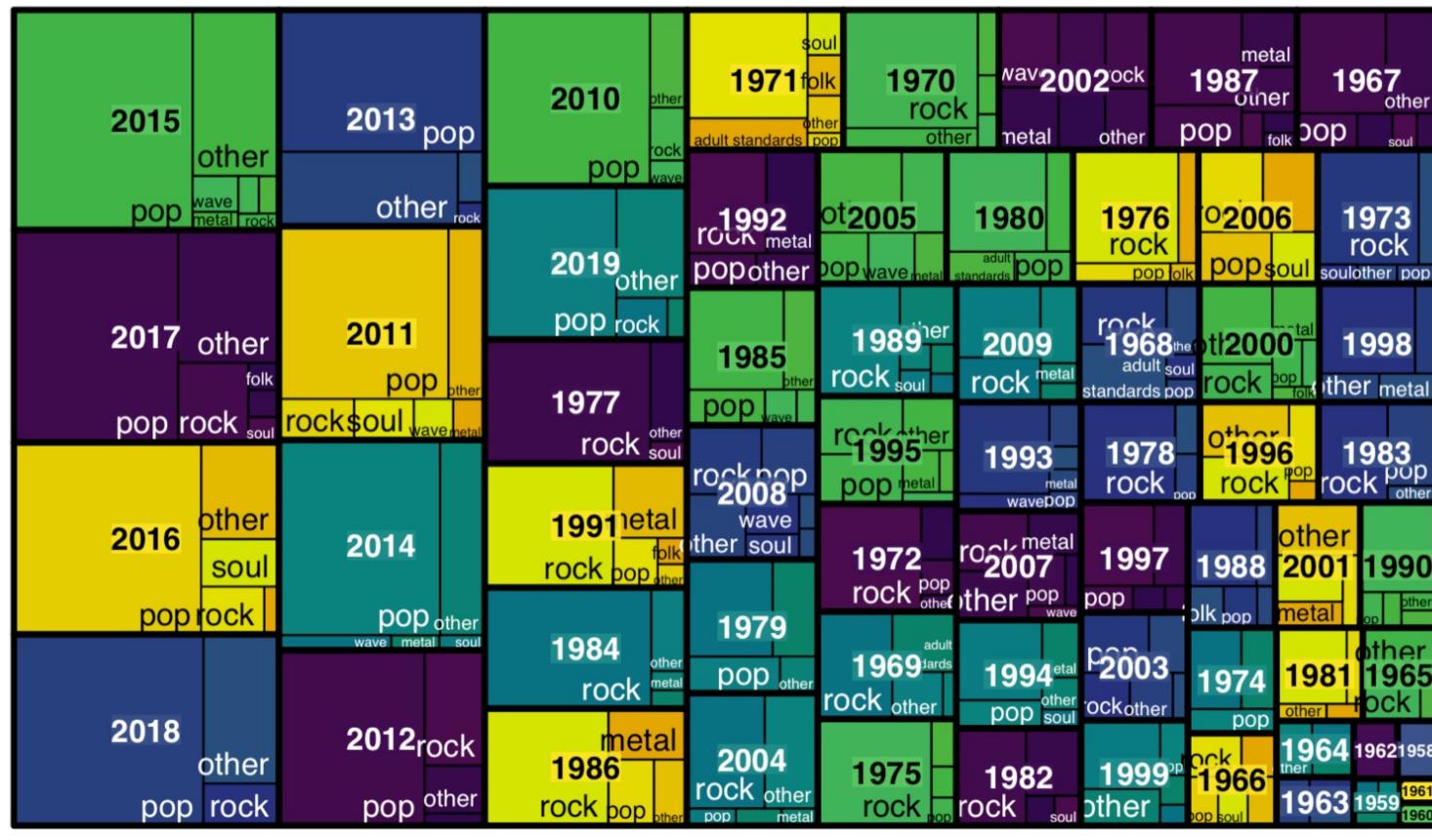


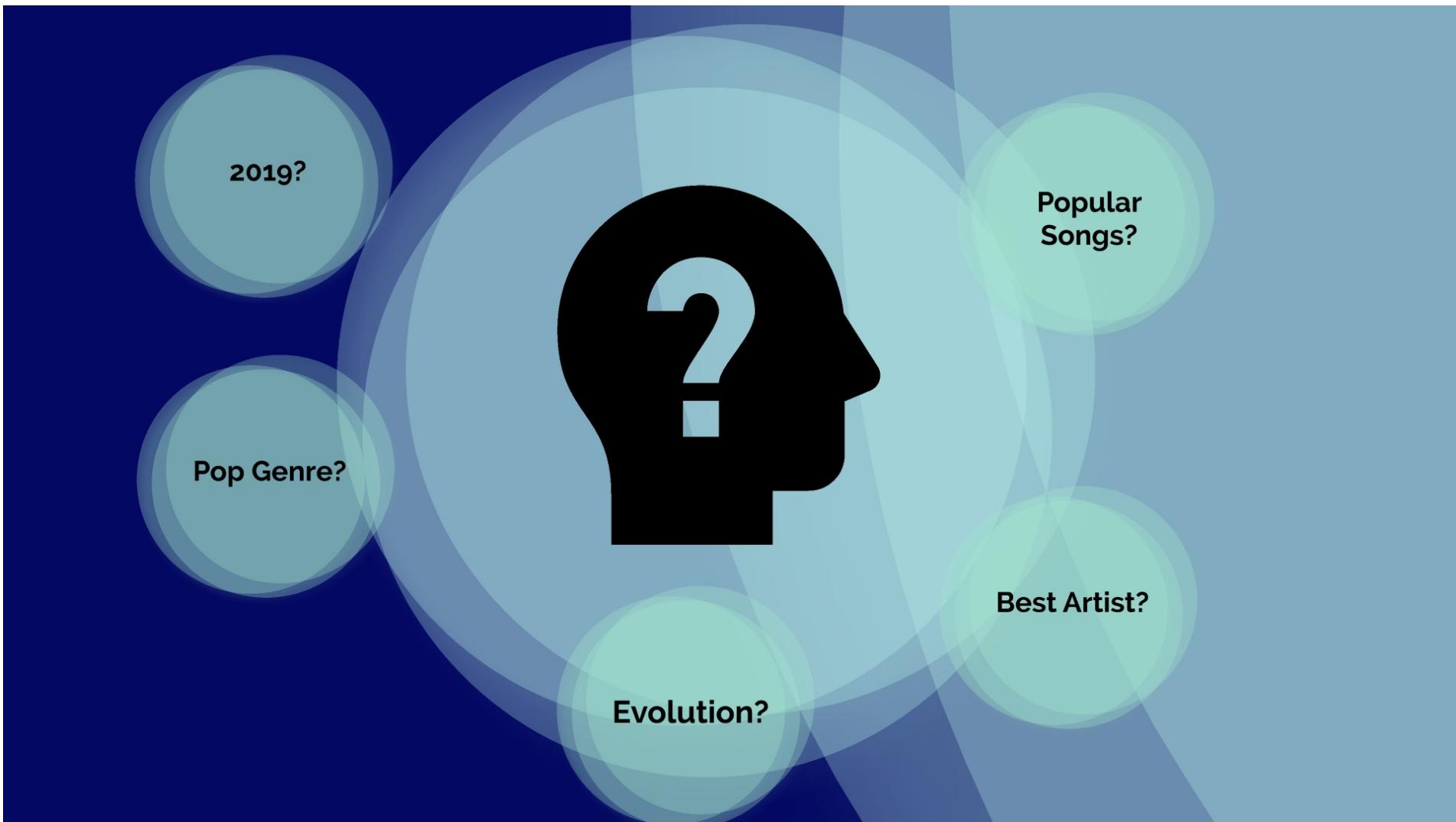






## Popularity





# Data Overview

If you don't  
know Statistics  
...

Numeric Features

Categorical Variables

# Spotify Music Popularity



Qingyin Ge  
qg2177

Data  
Description

Data  
Overview

Data  
Analysis



# Summary

- The average popularity of songs among different genres are different, pop, rock, metal, soul, and adult standards are the top 5 genres
- The average popularity of top songs in 2020 should be 72.85 based on arima model, and it should be 73.76 based on sample acf/pacf plot
- The final multivariate regression model selected regresses popularity on Year which is a two-factor variable containing value Centry20 and Centry21, and Energy, Danceability, Loudness, Liveliness, Acousticness, Speechiness. However, the result seems to be not satisfying, with adjusted multiple  $R^2 = 11\%$ . So we may claim that the mean popularity is not quite associated with the numeric features.

Which genre  
is different?

What's the  
popularity  
next year?

Do they affect  
music  
popularity?

Genres	Number	Average	SD
Pop	667	62.09	15.59
Rock	674	60.70	12.71
Soul	32	67.28	7.20
Metal	72	66.79	10.09
Adult Standards	87	62.75	9.50
Folk	15	63.87	7.99
Wave	30	66.33	11.66
Other	423	56.36	16.84

ANOVA  
p-value =  
 $3.4 \times 10^{-11}$

Mean of popularity is **different** among all the genres.

Bonferroni Confidence Interval is (3.7,18.1).



## Which ones are different?

### Tukey's HSD Procedure

	diff	lwr	upr	p adj
other-adult standards	-6.385424	-11.520200	-1.250649	$4.120544e-03$
other-rock	-4.334144	-7.039780	-1.628508	$3.459355e-05$
metal-other	10.429965	4.869228	15.990702	$4.042590e-07$
pop-other	5.729752	3.018647	8.440857	$5.032622e-09$
wave-other	9.971631	1.730545	18.212718	$6.041363e-03$
soul-other	10.919548	2.922555	18.916541	$9.321470e-04$

### Paired t-test with Holm's adj.

pop > other,  $5.0 \times 10^{-9}$   
 rock > other,  $3.3 \times 10^{-5}$   
 soul > other,  $8.9 \times 10^{-4}$   
 metal > other,  $3.9 \times 10^{-7}$   
 adult > other,  $4 \times 10^{-3}$

# Summary

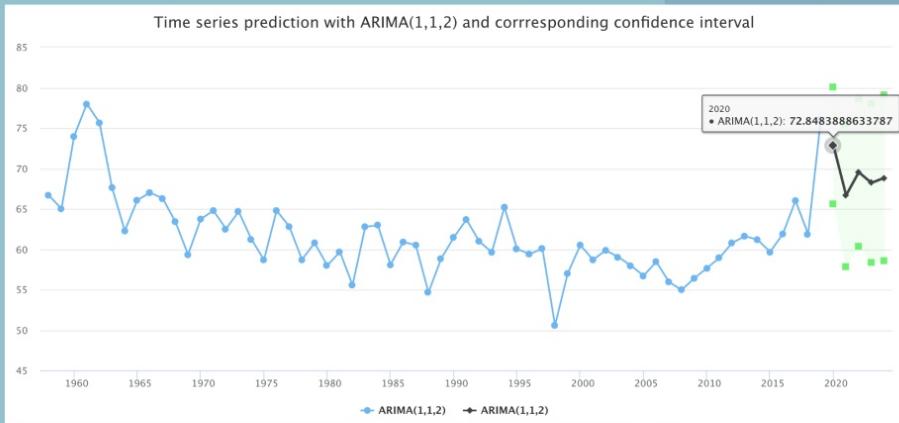
- The average popularity of songs among different genres are different, pop, rock, metal, soul, and adult standards are the top 5 genres
- The average popularity of top songs in 2020 should be 72.85 based on arima model, and it should be 73.76 based on sample acf/pacf plot
- The final multivariate regression model selected regresses popularity on Year which is a two-factor variable containing value Centry20 and Centry21, and Energy, Danceability, Loudness, Liveliness, Acousticness, Speechiness. However, the result seems to be not satisfying, with adjusted multiple  $R^2 = 11\%$ . So we may claim that the mean popularity is not quite associated with the numeric features.

Which genre  
is different?

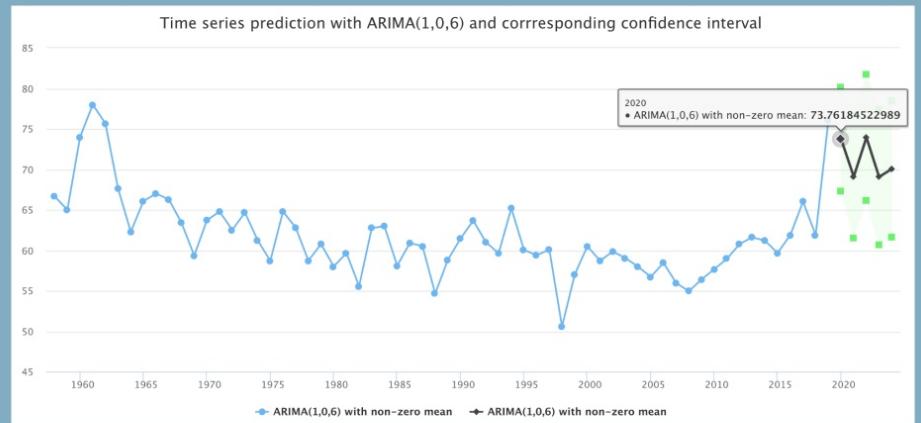
What's the  
popularity  
next year?

Do they affect  
music  
popularity?

auto.arima  
ARIMA(1,1,2)



Sample acf/pacf plot  
ARIMA(1,0,6)



# Summary

- The average popularity of songs among different genres are different, pop, rock, metal, soul, and adult standards are the top 5 genres
- The average popularity of top songs in 2020 should be 72.85 based on arima model, and it should be 73.76 based on sample acf/pacf plot
- The final multivariate regression model selected regresses popularity on Year which is a two-factor variable containing value Centry20 and Centry21, and Energy, Danceability, Loudness, Liveliness, Acousticness, Speechiness. However, the result seems to be not satisfying, with adjusted multiple  $R^2 = 11\%$ . So we may claim that the mean popularity is not quite associated with the numeric features.

Which genre  
is different?

What's the  
popularity  
next year?

Do they affect  
music  
popularity?



$$\begin{aligned}\text{Popularity} = & 77.42 - 5.60\text{Centrality} + 0.13\text{Energy} + 0.14\text{Danceability} \\ & + 1.49\text{Loudness} - 0.09\text{liveliness} - 0.03\text{Acousticness} + 0.26\text{Speechiness}\end{aligned}$$

Adjusted R<sup>2</sup> is 11%. By fitting polynomial regression model , using loess method and kernel smoothing, the result is not improved significantly. One possible reason is that the music popularity are **NOT** depend on those numeric features. People wouldn't choose to like one song by its loudness or liveliness. The characteristic of music that can attract people might be the rhythm, lyric or the singers, different people have different opinions on what is a mellifluous song.

## Summary

- The average popularity of songs among different genres are different, pop, rock, metal, soul, and adult standards are the top 5 genres
- The average popularity of top songs in 2020 should be 72.85 based on arima model, and it should be 73.76 based on sample acf/pacf plot
- The final multivariate regression model selected regresses popularity on Year which is a two-factor variable containing value Centry20 and Centry21, and Energy, Danceability, Loudness, Liveliness, Acousticness, Speechiness. However, the result seems to be not satisfying, with adjusted multiple  $R^2 = 11\%$ . So we may claim that the mean popularity is not quite associated with the numeric features.

Which genre  
is different?

What's the  
popularity  
next year?

Do they affect  
music  
popularity?

# Spotify Music Popularity



Qingyin Ge  
qg2177

Data  
Description

Data  
Overview

Data  
Analysis

