



# Relationship Between the Type of Music and Economic Conditions

Yoon Hye Sun  
Yu-Ting Cheng  
Chanju Oh

# Agenda

1. Introduction to dataset
2. Dictionary
3. Hypothesis
4. Background
5. Simplify complex categorical variables
6. Missing values
7. Handling Outliers
8. Split the dataset
9. Apply to a model
10. Interpretation
11. Business values & Recommendations
12. Reference

# Introduction to dataset

## Top Hits Spotify from 2000-2019

Top songs spotify playlists



[Data Card](#) [Code \(56\)](#) [Discussion \(6\)](#)

### About Dataset

#### Context

This dataset contains audio statistics of the top 2000 tracks on Spotify from 2000-2019. The data contains about 18 columns each describing the track and its qualities.

#### Content

- artist: Name of the Artist.
- song: Name of the Track.
- duration\_ms: Duration of the track in milliseconds.
- explicit: The lyrics or content of a song or a music video contain one or more of the criteria which could be considered offensive or unsuitable for children.
- year: Release Year of the track.

#### Usability ⓘ

10.00

#### License

Other (specified in description)

#### Expected update frequency

Never

#### Tags

Music

Multiclass Classification

# Dictionary

Target variable:

- popularity: The higher the value the more popular the song is.

Predictor variables:

- duration\_ms: Duration of the track in milliseconds.
- explicit: The lyrics or content of a song or a music video contain one or more of the criteria which could be considered offensive or unsuitable for children.
- year: Release Year of the track.
- 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.
- energy: Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity.
- key: The key the track is in. Integers map to pitches using standard Pitch Class notation. E.g. 0 = C, 1 = C#/D ♭, 2 = D, and so on. If no key was detected, the value is -1.
- loudness: The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typically range between -60 and 0 db.

# Dictionary

- mode: Mode indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. Major is represented by 1 and minor is 0.
- speechiness: Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.
- acoustictness: A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
- liveness: Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.
- valence: A measure from 0.0 to 1.0 describing 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).
- tempo: The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
- gdp: GDP is the total value of all goods and services produced within a country in a given time, showing the size and health of the economy. It includes consumer spending, business investments, government spending, and net exports (exports minus imports). It's a vital measure of economic activity and growth.
- happiness\_index: Positive emotional state of contentment, joy, and well-being, influenced by relationships, personal fulfillment, health, and life circumstances. Desired for overall quality of life.
- unemployment\_rate: Percentage of jobless individuals actively seeking work within the labor force, indicating job market health and economic conditions. Calculated by dividing unemployed by labor force, multiplied by 100.

# Hypothesis

“During times of economic downturn, we suppose there is an increase in the popularity of upbeat and energetic genres such as electronic dance music, while during economic growth periods, more soulful, introspective, thoughtful music, tend to dominate the charts.”

Another recent trend influenced by the economy is the mainstreaming of electronica-style dance music, says Alexandra Smith, director of consumer trends at market research company, [Intel](#). She points to bands like [Skrillex](#), [Lady Gaga](#) and [Robyn](#).

“It’s put-your-headphones-on, tune-the-world-out music,” she says. “It tends to be really upbeat. And it really makes sense, when times are tough, you’re going to want to turn to that type of music.”

[For pop music, it's the economy, stupid! - Marketplace](#)

# Background

2006-2007 (Pre-Recession): During this period of economy growth and stability, with lowest tempo accounted for more percentage than the other periods.

2008-2009 (Global Financial Crisis): The global financial crisis hit, positive songs were more popular.

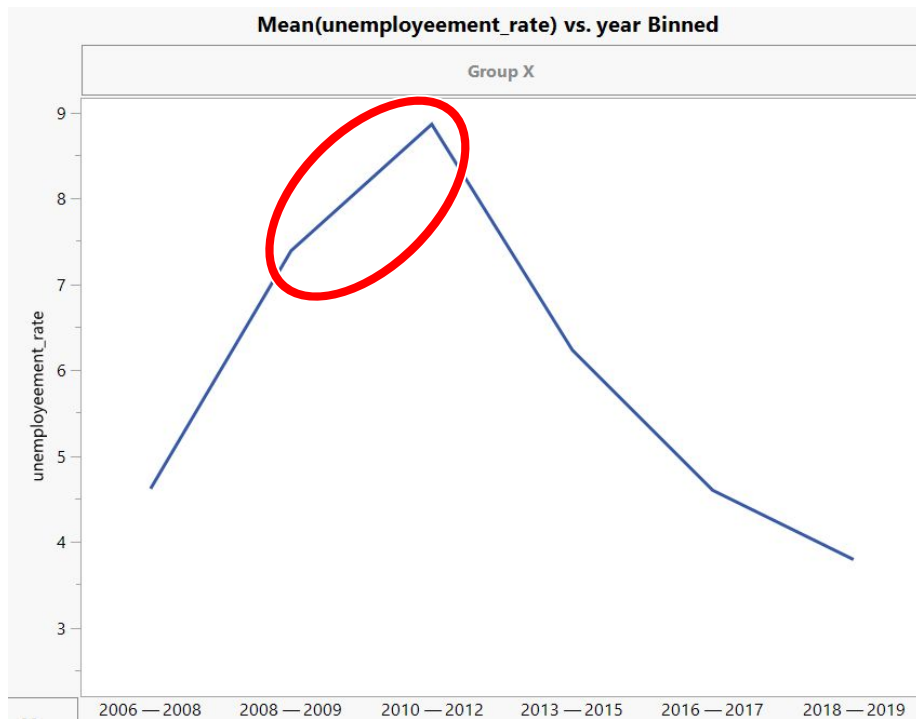
2010-2012 (Post-Recession Recovery): The economy began recovering, but the unemployment rate was the highest, songs with higher tempo accounted for more percentage than the other periods.

2013-2015 (Steady Recovery): With economy stabilizing, a diverse range of genres gained popularity, more hip hop songs were getting popular. And negative songs accounted for more percentage than the other periods.

2016-2017 (Continued Growth): As economy continued to grow, music trends diversified further. Hip hop songs got more popular and negative songs accounted for more percentage than most of the other periods.

2018-2019 (Late-stage Growth): The economy had already experienced substantial growth, and it was approaching a more mature and stable state. In this period, the music scene continued to embrace a mix of genres. More latin and R&B songs were in top hit songs.

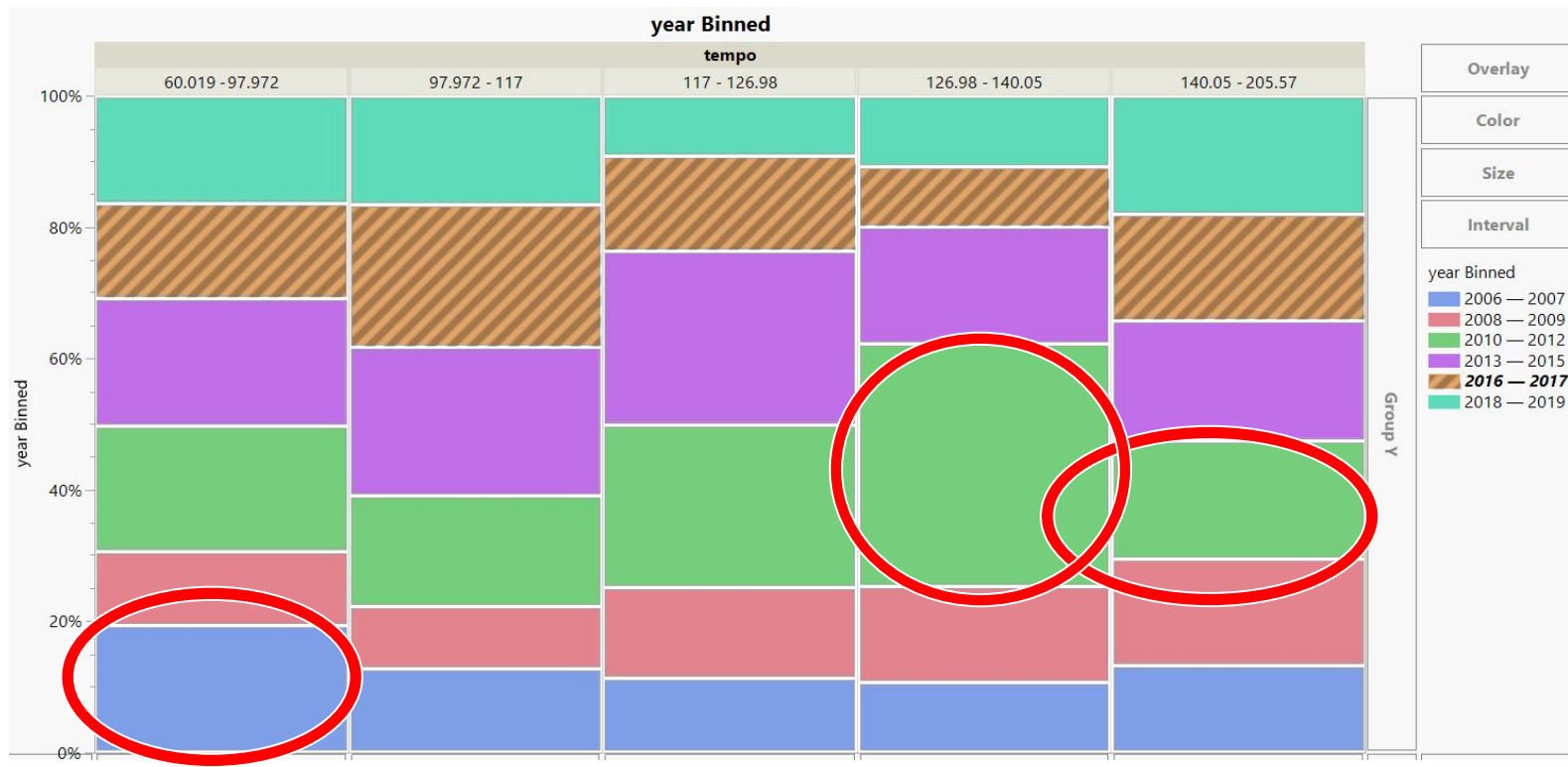
# Background



Unemployment rate getting higher during 2008-2012



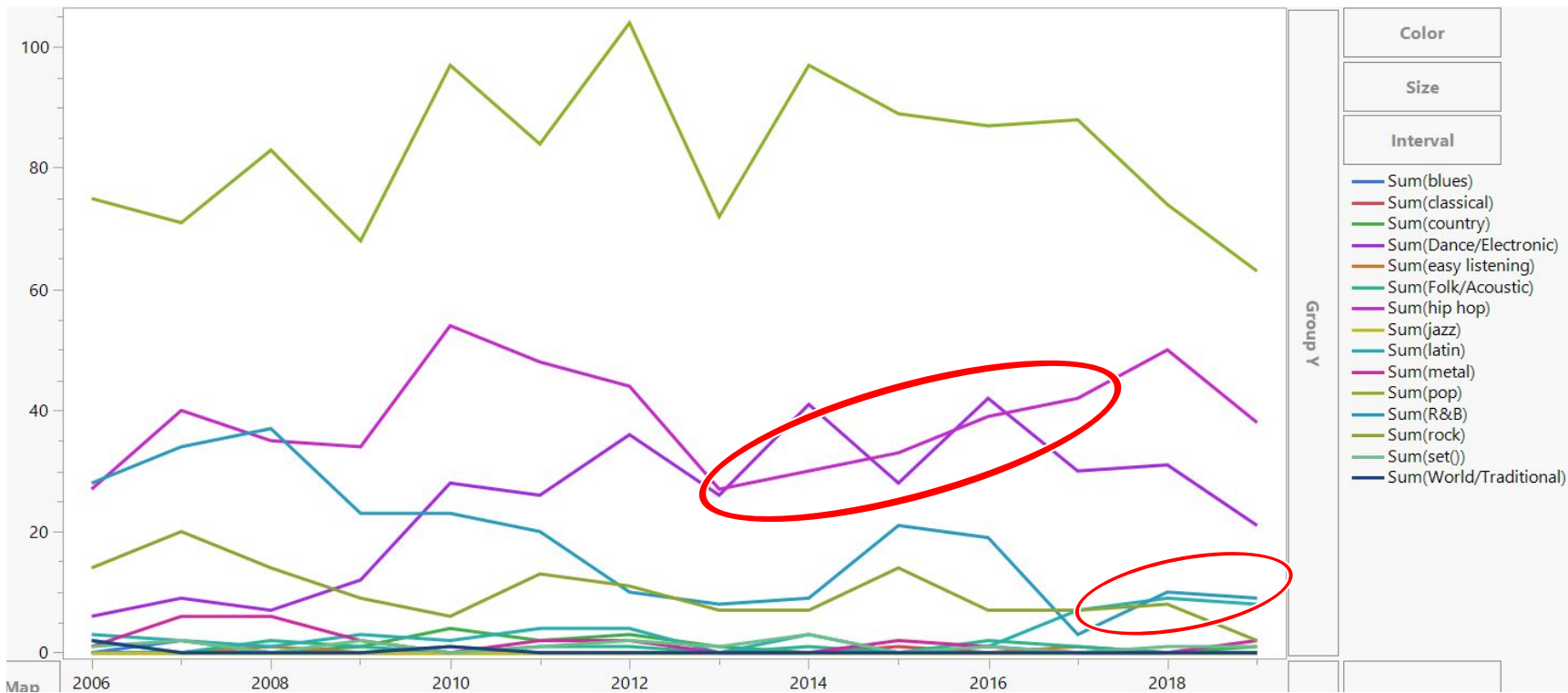
# Background



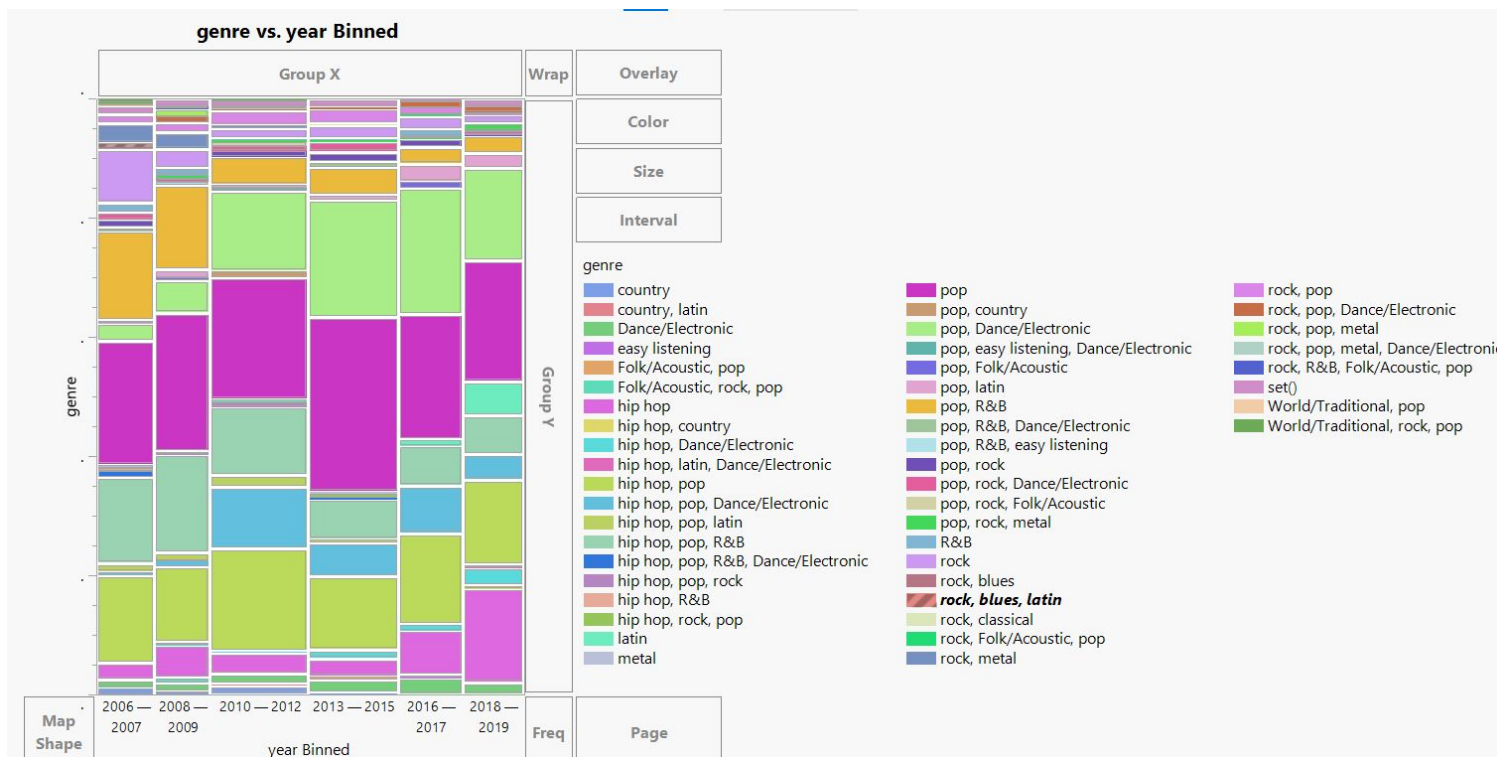
# Background



# Background

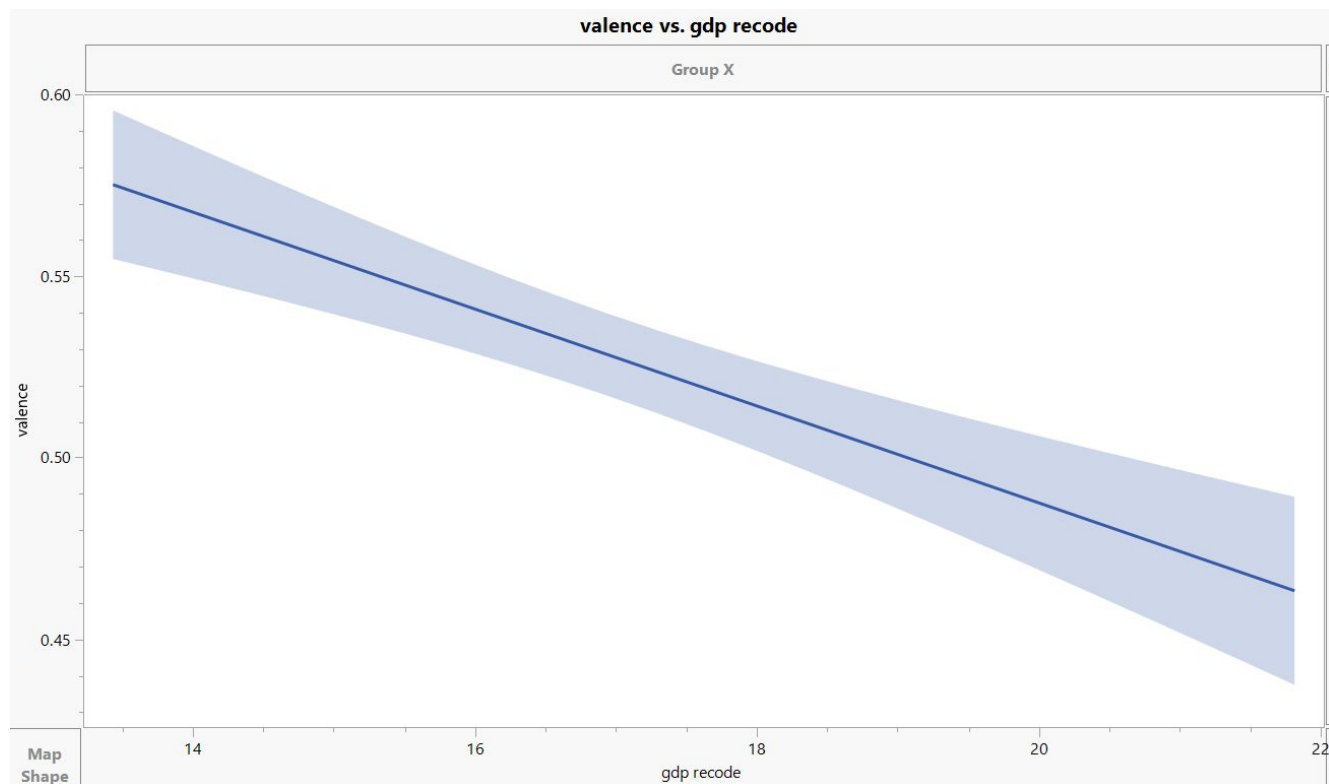


# Background



More different genres are popular when the economic is or getting better

# Background



People would like to listen to positive songs, when the economic is bad

# Simplify complex categorical variables

- Problem:  
Genre column is categorical but is not properly formatted -> transformation is needed
- Solution:  
Split up genres and convert each genre into a binary variable

genre							
	blues	classical	country	Dance/ Electronic	easy listening	Folk/Acoustic	hip hop
pop							
rock, pop							
pop, country							
rock, metal							
pop							
hip hop, pop, R&B							
hip hop							
pop, rock							
pop, R&B							
Dance/Electronic							
pop							
pop							
pop							
pop							
rock, pop							
pop, Dance/Electronic							
hip hop, pop, R&B							
pop							
pop							
pop							
hip hop							
rock, metal							
rock, Folk/Acoustic, easy listening							

# Missing values

- Exclude missing values

Column	Number Missing
gdp recode	42
happiness_index	611

	Rows
All rows	2,000
Selected	1
Excluded	611
Hidden	0
Labeled	0

**Explore Outliers**

**Quantile Range Outliers**

Some quantiles were stretched to avoid a large group at the median.  
Some tail quantiles were no different from the median.

Outliers are values Q times the interquantile range past the lower and upper quantiles.

Tail Quantile: 0.1   
Q: 3 ☐ Restrict search to integers

Outliers by Column ☒ Outliers by Cell ☐ Outliers by Row

☐ Show only columns with outliers

Select columns and choose an action.

Identify Outliers in Table:

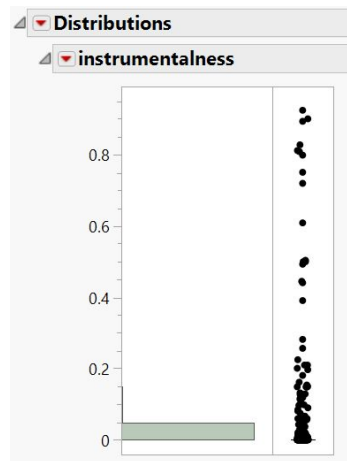
Clear Outliers in Table:

Column	Lower Prob	Upper Prob	Lower Quantile	Upper Quantile	Low Threshold	High Threshold	Number of Outliers	Outliers (Count)
duration_ms	0.1	0.9	181826	269373	-80815	532014	0	
year	0.1	0.9	2007	2018	1974	2051	0	
gdp recode	0.1	0.9	14.4489	20.6119	-4.0401	39.1009	0	
happiness_index	0.1	0.9	6.86395	7.28039	5.61463	8.5297	0	
popularity	0.1	0.9	1	79	-233	313	0	
danceability	0.1	0.9	0.472	0.835	-0.617	1.924	0	
energy	0.1	0.9	0.516	0.905	-0.651	2.072	0	
key	0.1	0.9	0	11	-33	44	0	
loudness	0.1	0.9	-7.67	-3.293	-20.801	9.838	0	
mode	0.1	0.9	0	1	-3	4	0	
speechiness	0.1	0.9	0.0322	0.246	-0.6092	0.8874	0	
acousticness	0.1	0.9	0.00245	0.348	-1.0342	1.38465	0	
instrumentalness	0.1	0.9	0	0.02254	0.0012	0.01192	88	0.0126 0.0127(2) 0.0128 0.0139 0.014
liveness	0.1	0.9	0.0679	0.355	-0.7934	1.2163	0	
valence	0.1	0.9	0.234	0.812	-1.5	2.546	0	
tempo	0.1	0.9	90.007	155.974	-107.89	353.875	0	
classical	0.1	0.9992	0	0.91406	-2.7422	3.65625	0	
country	0.1	0.9938	0	1	-3	4	0	
Dance/Electronic	0.1	0.9	0	1	-3	4	0	
easy listening	0.1	0.9984	0	1	-3	4	0	
Folk/Acoustic	0.1	0.9938	0	1	-3	4	0	
hip-hop	0.1	0.9	0	1	-3	4	0	
jazz	0.1	0.9	0	0	-1e-70	1e-70	0	
latin	0.1	0.975	0	1	-3	4	0	
metal	0.1	0.9875	0	1	-3	4	0	
pop	0.1	0.9	0	1	-3	4	0	
R&B	0.1	0.9	0	1	-3	4	0	
rock	0.1	0.9	0	1	-3	4	0	
World/Traditional	0.1	0.9984	0	1	-3	4	0	

# Handling Outliers

- Outlier Report indicates that only “instrumentalness” variable has 86 outliers
- Normalize the distribution of instrumentalness

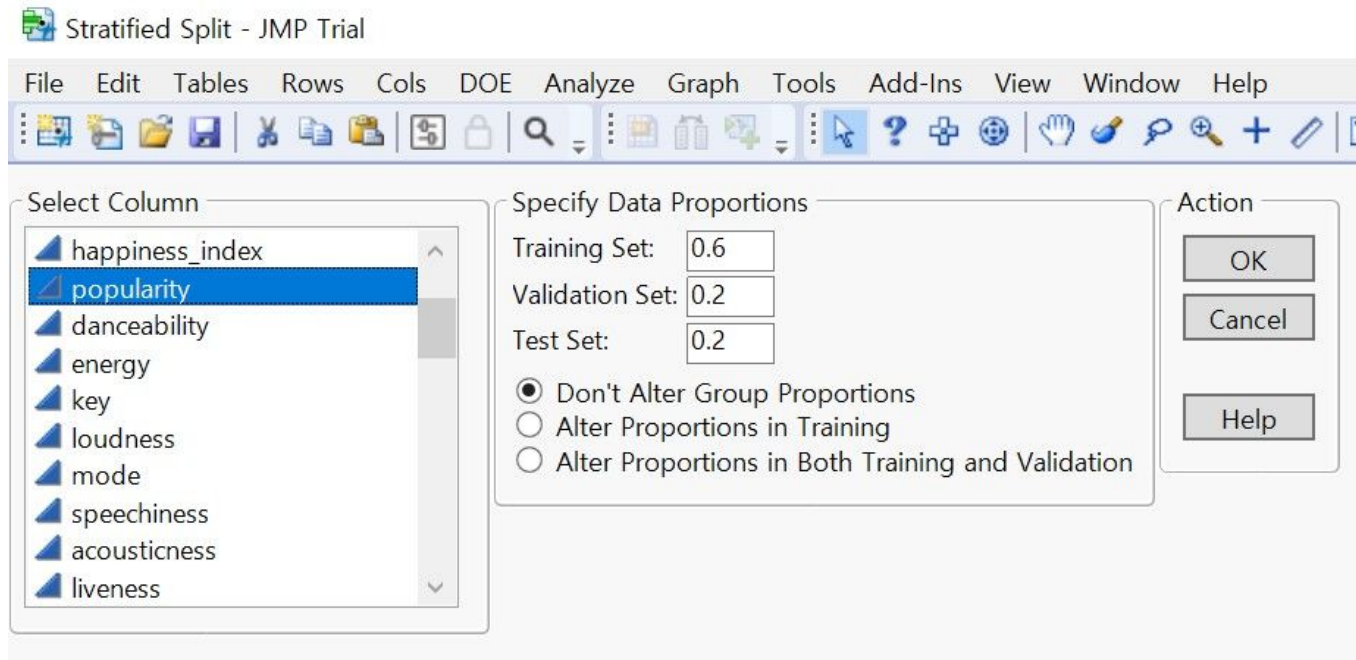
Column	Lower Prob	Upper Prob	Lower Quantile	Upper Quantile	Low Threshold	High Threshold	Number of Outliers	Outliers (Count)
duration_ms	0.1	0.9	181826	269373	-80815	532014	0	0
year	0.1	0.9	2007	2018	1974	2051	0	0
gdp_recode	0.1	0.9	14.4489	20.6119	-4.0401	39.1009	0	0
happiness_index	0.1	0.9	6.86395	7.28039	5.61463	8.5297	0	0
popularity	0.1	0.9	1	79	-233	313	0	0
danceability	0.1	0.9	0.472	0.835	-0.617	1.924	0	0
energy	0.1	0.9	0.516	0.905	-0.651	2.072	0	0
key	0.1	0.9	0	11	-33	44	0	0
loudness	0.1	0.9	-7.67	-3.293	-20.801	9.838	0	0
mode	0.1	0.9	0	1	-3	4	0	0
speechiness	0.1	0.9	0.0322	0.246	-0.6092	0.8874	0	0
acousticness	0.1	0.9	0.00245	0.348	-1.0342	1.38465	0	0
instrumentalness	0.1	0.9	0	0.00264	-0.0079	0.01056	86	0.0126 0.0127(2) 0.0128 0.0139 0.01
liveness	0.1	0.9	0.0679	0.355	-0.7934	1.2163	0	0
valence	0.1	0.9	0.234	0.812	-1.5	2.546	0	0
tempo	0.1	0.9	90.007	155.974	-107.89	353.875	0	0
classical	0.1	0.9992	0	0.91406	-2.7422	3.65625	0	0
country	0.1	0.9938	0	1	-3	4	0	0
Dance/Electronic	0.1	0.9	0	1	-3	4	0	0
easy listening	0.1	0.9984	0	1	-3	4	0	0
Folk/Acoustic	0.1	0.9938	0	1	-3	4	0	0
hip hop	0.1	0.9	0	1	-3	4	0	0
jazz	0.1	0.9	0	0	-1e-70	1e-70	0	0
latin	0.1	0.975	0	1	-3	4	0	0
metal	0.1	0.9875	0	1	-3	4	0	0
pop	0.1	0.9	0	1	-3	4	0	0
R&B	0.1	0.9	0	1	-3	4	0	0
rock	0.1	0.9	0	1	-3	4	0	0
World/Traditional	0.1	0.9984	0	1	-3	4	0	0



SHASH Transform to Normal	
instrumentalness	
1.4380170818	
1.3660166114	
-1.103712012	
1.4183153287	
1.3146964916	
1.3301410041	
-1.103712012	
1.4077168365	
1.4565137522	
-1.103712012	
-1.103712012	
1.477879484	
-1.103712012	
-1.103712012	



# Split the dataset



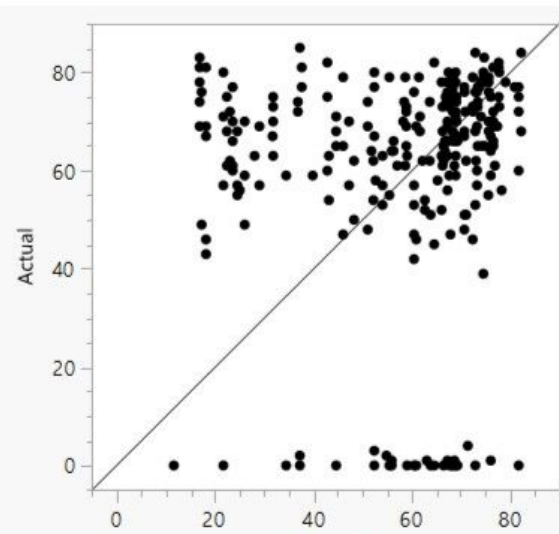
# Apply to a model

## Partition for popularity



	RSquare	RASE	N	Number of Splits	AICc
Training	0.548	16.141587	850	134	7464.84
Validation	-0.52	28.511491	273		
Test	-0.58	29.389816	266		

## Test Set



# Apply to a model

## Column Contributions



Term	Number of Splits	SS	Portion
duration_ms	9	31403.0065	0.1172
acousticness	9	25357.9627	0.0946
key	9	22565.4359	0.0842
loudness	9	20914.0613	0.0780
valence	9	19668.8041	0.0734
speechiness	11	19575.4834	0.0730
energy	9	18461.2322	0.0689
danceability	11	16652.158	0.0621
unemployment_rate	6	13906.0749	0.0519
liveness	6	13324.3944	0.0497
hip hop	4	10675.8745	0.0398
Dance/Electronic	1	9856.81715	0.0368
SHASH Transform to Normal instrumentalness	2	8406.00598	0.0314

inflation	6	8209.72952	0.0306
rock	2	6339.12645	0.0237
gdp recode	3	5282.51904	0.0197
mode	9	5018.66205	0.0187
pop	4	3248.93613	0.0121
explicit	2	2868.37155	0.0107
R&B	2	2189.50454	0.0082
tempo	7	2158.46786	0.0081
happiness_index	4	1909.86991	0.0071
year	0	0	0.0000
classical	0	0	0.0000
country	0	0	0.0000
easy listening	0	0	0.0000
Folk/Acoustic	0	0	0.0000
jazz	0	0	0.0000
latin	0	0	0.0000
metal	0	0	0.0000
World/Traditional	0	0	0.0000



# Interpretation

As the model result says, Unemployment Rate has the highest correlation among economic factors with target variable, Popularity. All other economic factors has positive correlation with target variable. As unemployment rate and inflation rate get higher, the songs that has high duration\_ms, acousticness, key, loudness, valence, speechiness, energy and danceability are popular.

# Business values & Recommendations

According to our study, we find out that the economy impact the popularity of different types of songs.

During the economic turndown, we recommend the music producers to make high duration\_ms, acousticness, key, loudness, valence, speechiness, energy and danceability music. Meanwhile, when economy is booming, we recommend the music producers to make low duration\_ms, acousticness, key, loudness, valence, speechiness, energy and danceability music.

# Reference

[For pop music, it's the economy, stupid! - Marketplace](#)

[Top Hits Spotify from 2000-2019 | Kaggle](#)

[GDP of each country and region\(1960-2020\) | Kaggle](#)

[U.S. Unemployment Rate 1991-2023 | MacroTrends](#)

[World Happiness Report, 2005-Present | Kaggle](#)

[USA Key Economic Indicators | Kaggle](#)