# Bustin Jieber’s next big break

…

BUS. 5100-93

Group 1: Wendy G.,Virginia F., Emily D., Teezal G., Ricardo E.

# Setting the stage

- We are managers of up-and-coming musician Bustin Jieber
- We've partnered with Spotify to review streaming data
- Our goal is to analyze data to provide our artist with recommendations on what type of song to release:
  - A **danceable** spicy and upbeat track
  - Slow, love song

# Data Source and Location of Analysis

**Location of Datasets and Story:** Cal State LA  folder > Virginiaf > Group 1 Data (folder)

**Story Location:** Cal State LA  folder > Virginiaf > Group 1 Data > Danceability Predictions by Year Story

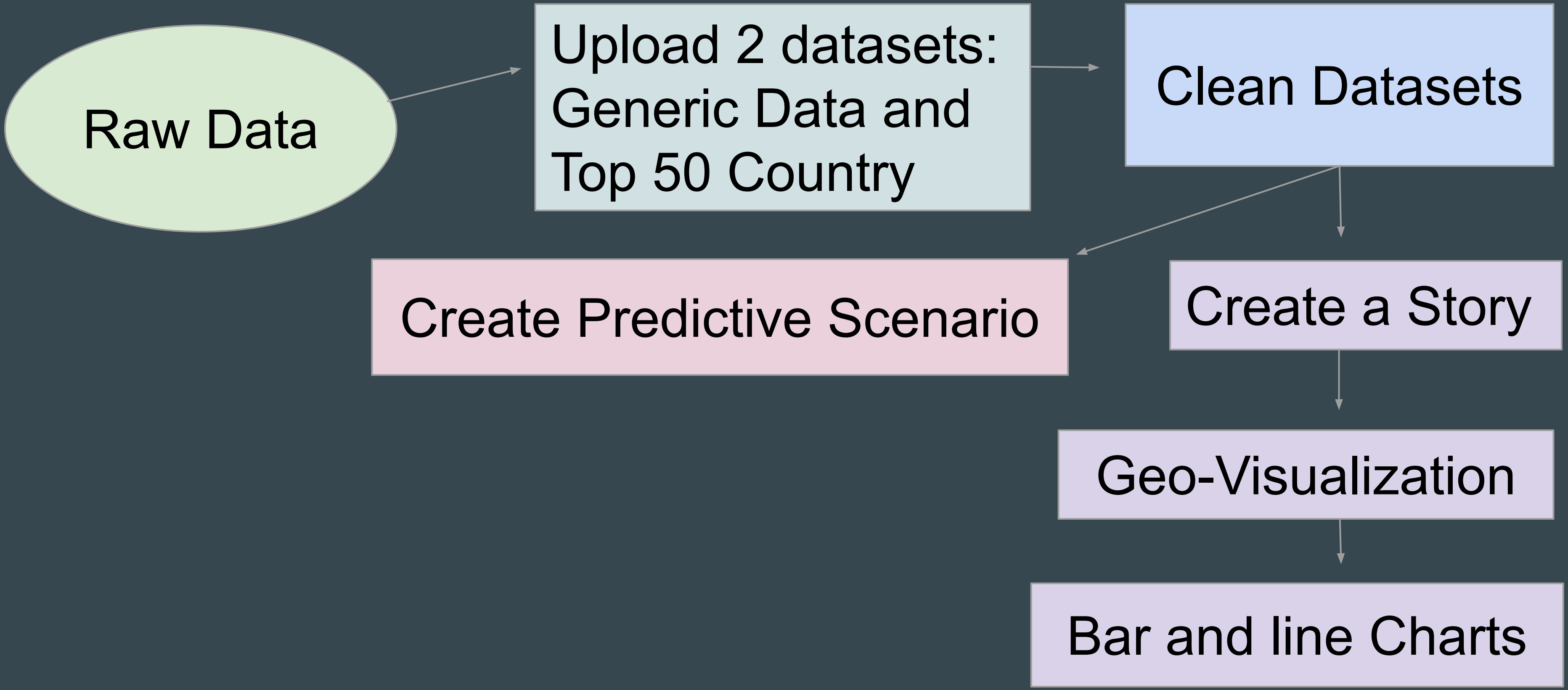
**Data Set Source:** Kaggle.com

**GitHub: vfrazee/SpotifyDataAnalysis**

*Table 1. Data Specification*

| Data Set | Size |
|---|---|
| Spotify Genetic Data | 33.1 MB |
| Top 50 | 118 KB |

# Flowchart


Raw Data
Upload 2 datasets: Generic Data and Top 50 Country
Clean Datasets
Create Predictive Scenario
Create a Story
Geo-Visualization
Bar and line Charts

# Using Data Analysis to Understand This Information

1. What influencers most affect danceability?
2. How can we validate this data?
3. Can we predict and visualize this information in a way that is easier to understand?
4. How can we use this information to choose our next single and our touring/marketing schedule?

# Understanding How Spotify Classifies Songs

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.

# Step 1: Regression Analysis

The goal of this Regression Analysis is to look at how different influencers affect danceability scores, so we know what kind of song to produce.

Then we can apply this information to predict how danceable popular songs will are year over year.

Predictive Model 1: Set influencers to all
Predictive Model 2: Limited influencers to just 5

# Model Results

## Model 1

| Influencer | Contribution |
|---|---|
| valence | 24.93% |
| tempo | 15.69% |
| speechiness | 13.26% |
| energy | 11.71% |
| Release Year | 9.81% |

## Model 2

| Influencer | Contribution |
|---|---|
| valence | 31.86% |
| tempo | 20.64% |
| year | 18.79% |
| speechiness | 17.48% |
| energy | 11.23% |

# Regression Analysis Results

Predictive Models (2)

| | Name | Status | Creation Date | Root Mean Square Error (RMSE) | Prediction Confidence |
|---|---|---|---|---|---|
| ☑ | Model 2<br>Spotify Predictive Mode... | Trained | Mar 9, 2021 19:19:26 | 0.108 | 97.81% |
| ☑ | Model 1<br>Spotify Regressive Anal... | Trained | Mar 9, 2021 19:01:50 | 0.098 | 98.17% |

In both Models, we can see that Valance is the highest influencer (24.93% and 31.86% ).

The accuracy of the predictions are 98.17% and 97.81%.

# Apply the Predictive Model

Using Model 1, we created a new Predictive Column to our dataset that predicts Danceability by year on a 0-1 scale.

| danceability | energy | liveness | loudness | popularity | tempo | valence | Release Y... | Predicted... |
|---|---|---|---|---|---|---|---|---|
| 0.397 | 0.252 | 0.113 | -16.669 | 17 | 101.625 | 0.662 | 1963 | 0.552402436733 |
| 0.467 | 0.0440000000000 | 0.1040000000000 | -25.111 | 28 | 67.124 | 0.502 | 1966 | 0.441204845905 |
| 0.742 | 0.8140000000000 | 0.1040000000000 | -2.677 | 29 | 117.1870000000 | 0.855 | 1968 | 0.603301048278 |
| 0.37 | 0.785999999999 | 0.113 | -8.866 | 20 | 139.306 | 0.273999999999 | 1968 | 0.325611114501 |
| 0.652 | 0.16 | 0.1040000000000 | -15.06299999999 | 21 | 96.477 | 0.457999999999 | 1969 | 0.508549094200 |
| 0.616 | 0.5870000000000 | 0.100999999999 | -11.297 | 30 | 186.52 | 0.905999999999 | 1973 | 0.459855020046 |
| 0.703 | 0.762 | 0.1040000000000 | -10.25 | 31 | 107.844 | 0.966 | 1973 | 0.683929085731 |
| 0.711 | 0.228999999999 | 0.113 | -17.813 | 26 | 115.855 | 0.376 | 1974 | 0.534811496734 |
| 0.488 | 0.0992 | 0.113 | -14.87900000000 | 26 | 134.4560000000 | 0.532999999999 | 1974 | 0.509336948394 |
| 0.650999999999 | 0.785999999999 | 0.1040000000000 | -8.436 | 34 | 135.041 | 0.392999999999 | 1979 | 0.455187261104 |

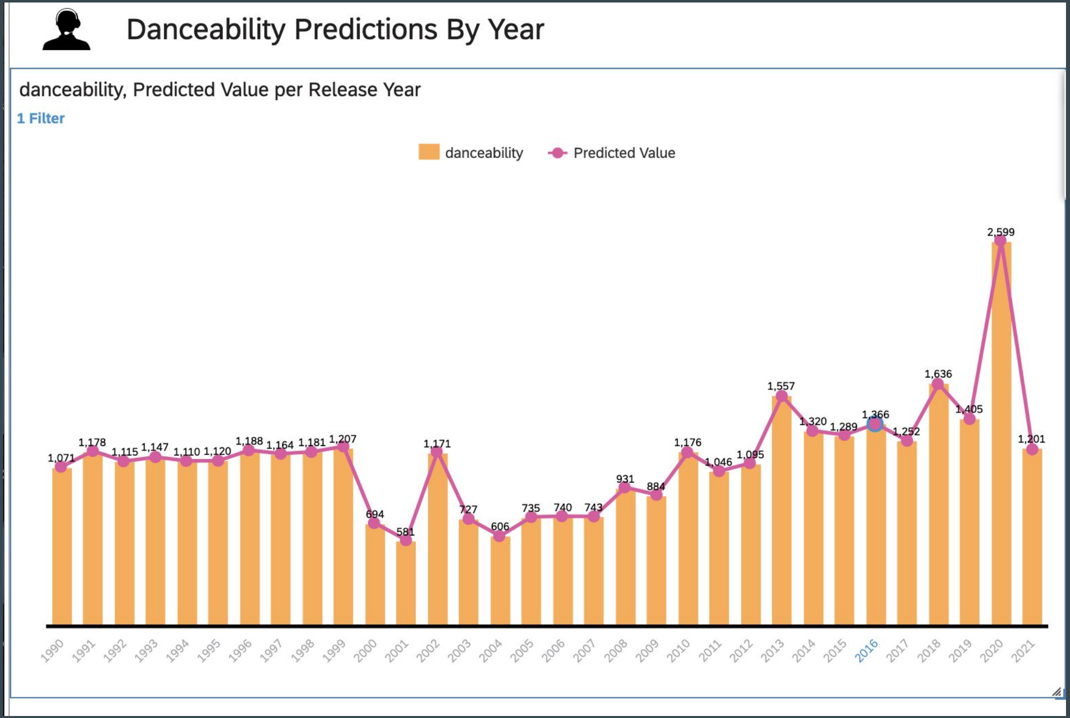
# Creating A Visual Story From Our Analysis

Now, we want to create a visual representation that demonstrates the accuracy of our predictions and validates the importance of the influencers we have identified as the most important for determining Danceability using a Story in SAP SAC.

Comparison Combination Column and Line chart that differentiates between the actual danceability and the Predicted Value.

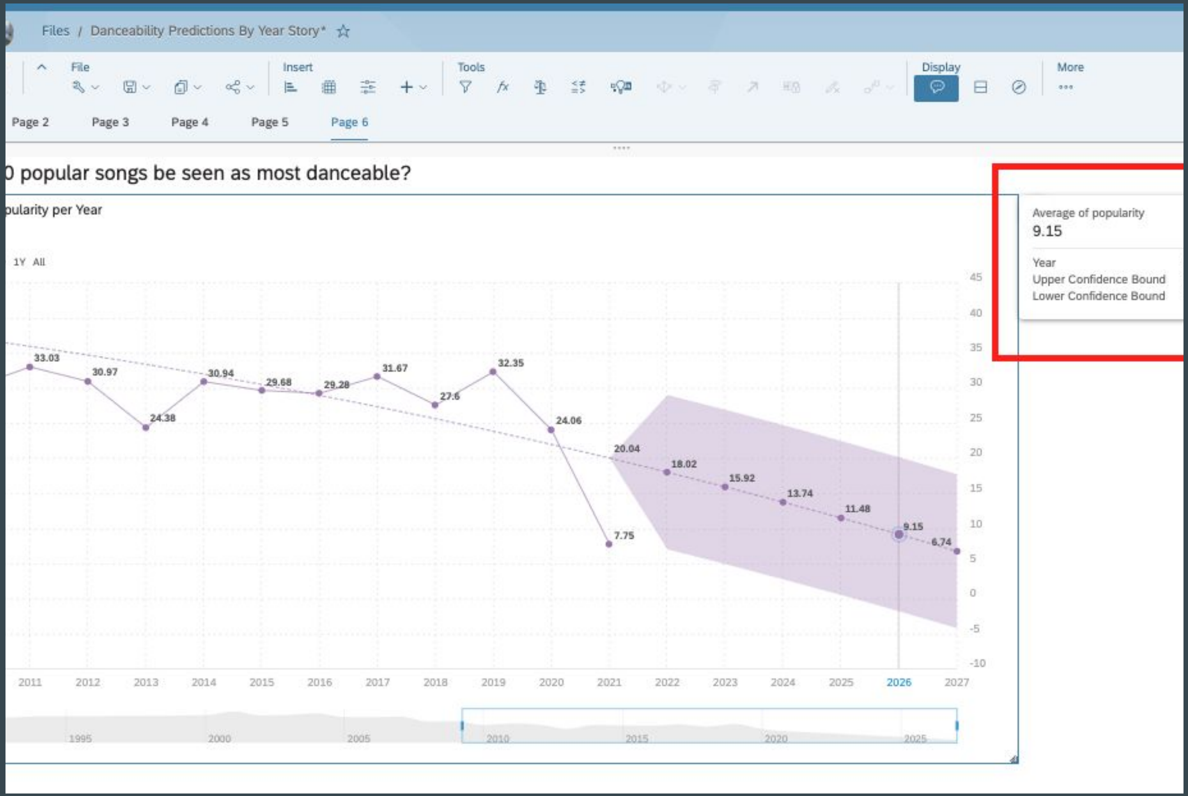
Segment the Dimensions by Release Year.

# Results




This exercise taught us a
4 key things

# Time Series - Popularity of songs by 2025




Time Series Analysis to understand the impact of popularity/danceability and top charts.

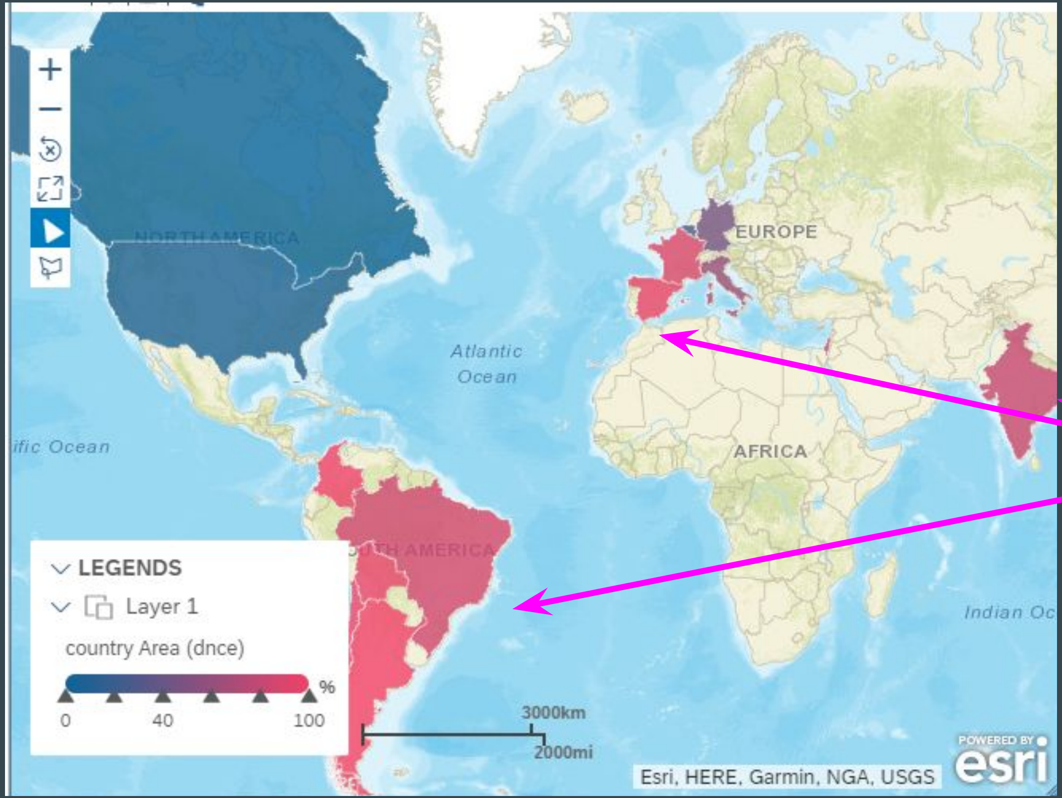
The average popularity score of songs will continue to decrease.

# Time Series - Danceability of songs by 2025


Time series analysis of Danceability to forecast whether the scores will continue to increase.

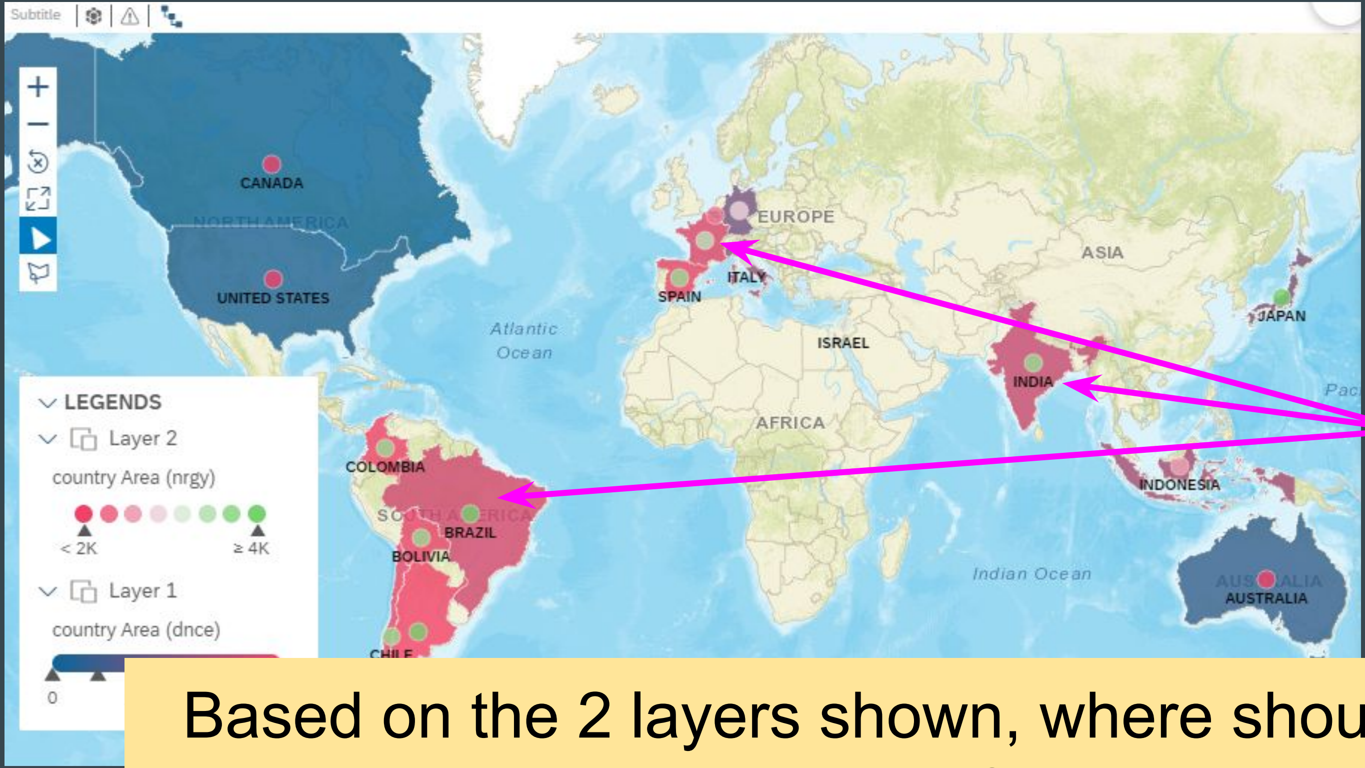
The chart indicates, using triple exponential smoothing, that danceability will continue to increase into 2025.

# Geo Mapping to target releases




Layer 1 - Choropleth drill colors illustrates countries with most danceability.

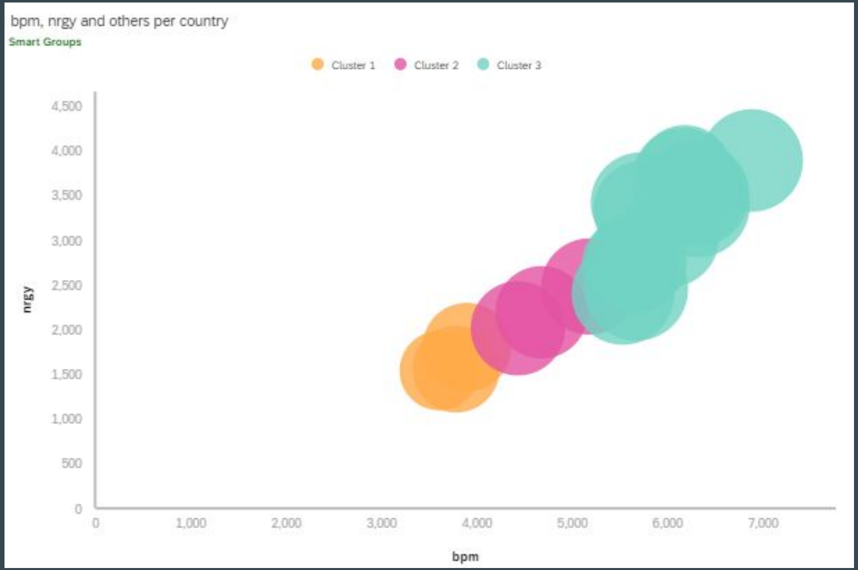
# Geo Mapping to target releases (cont.)

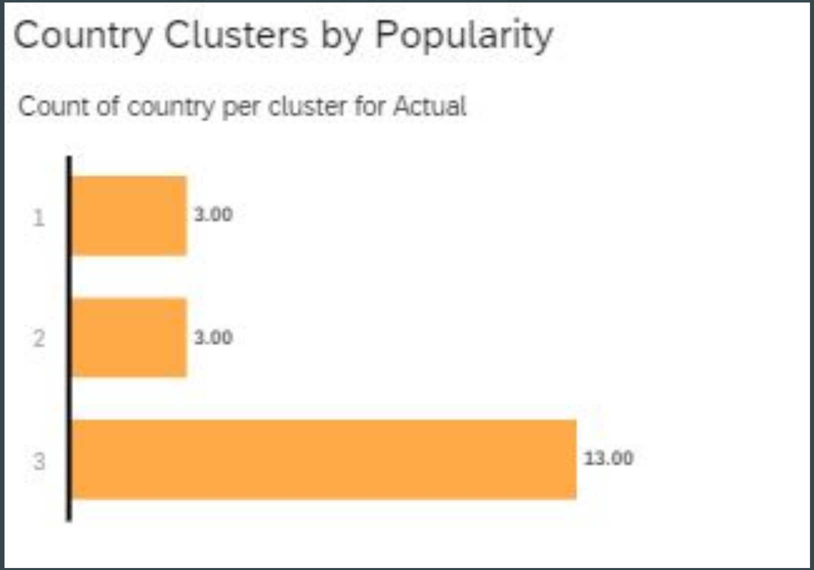



Layer 2 - bubbles indicate countries with songs that has most energy.

Based on the 2 layers shown, where should Bustin release a song with lots of rhythm and beats?

# Clustering Countries to segment based on shared attributes



# Correlation between bpm, nrgy and popularity of a song per country

# Conclusion

- As Managers of Bustin Jieber, we recommend an upbeat and energetic song..
- Construction of our song needs to leverage the forecasted danceability as indicated in our time series analysis.
- We need to release and tour a song in South America, Malaysia, Europe and India