Salsa/Bachata Classifier

Marco Sanchez-Ayala
Flatiron School

What are they?

- Musical genres that stem from afro-caribbean music

Can we distinguish between them?

Salsa

- Tempo: 160 - 220 bpm
- Percussion: conga-heavy
- Varying form (song structure)
- Harmonic progressions by 5ths (somewhat similar to jazz)

Bachata

- Tempo: 90 - 150 bpm
- Percussion: bongos/guira
- Simple, predictable form
- Simple, often four chord harmonic progressions

Can we distinguish between them?

Salsa

- Tempo: 160 - 220 bpm
- Percussion: conga-heavy
- Varying form (song structure)
- Harmonic progressions by 5ths (somewhat similar to jazz)

Bachata

- Tempo: 90 - 150 bpm
- Percussion: bongos/guira
- Simple, predictable form
- Simple, often four chord harmonic progressions

These differences may not be as obvious to the untrained ear!

Can we train a machine to distinguish between them?

Yes ...

Model	Accuracy Score (%)	Precision Score	Recall Score	F1 Score
Random Forest No engineered features	93.9	0.961	0.905	0.932

... with 94% accuracy!

Data Source: Spotify

- 1400 songs from 8 artists (4 salsa, 4 bachata)
- song ratio: 45:55 salsa vs bachata

Audio Features

- Duration
- Tempo
- Key
- Energy
- Danceability
- Instrumentalness
- etc.

Audio Analysis

- Start time and duration of each musical section


Data Source: Spotify®

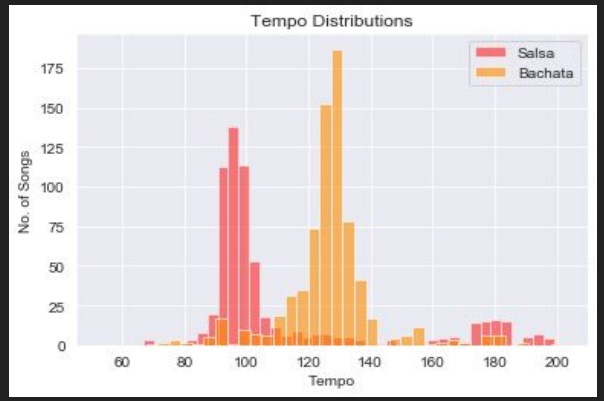
Audio Analysis

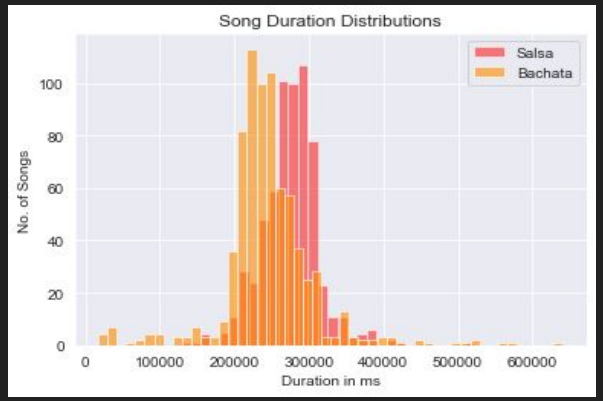
- Start time and duration of each musical phrase

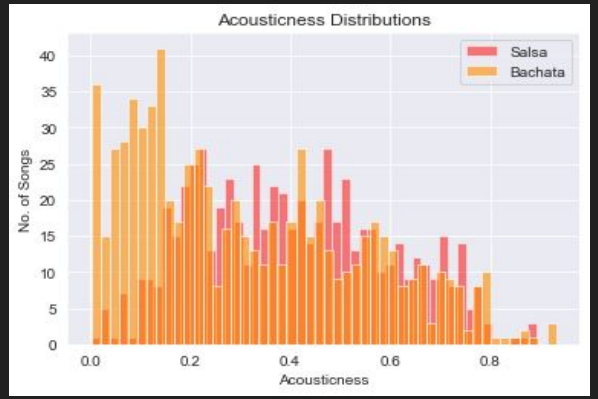
Engineered Features

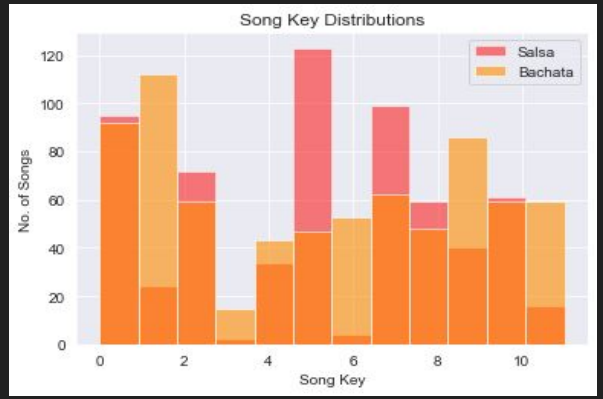
- No. of musical sections
- Avg. length of musical sections

Some Musical Features









Comparing Model Performance

Based on F1 score - false positives and false negatives are equally bad

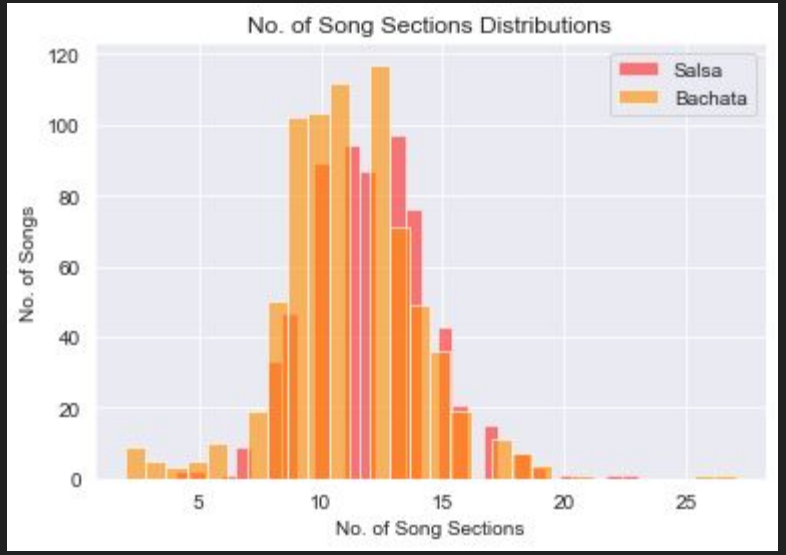
Model	Accuracy Score (%)	Precision Score	Recall Score	F1 Score
Decision Tree No engineered features	92.0	0.938	0.884	0.910
Random Forest No engineered features	93.9	0.961	0.905	0.932
Decision Tree With engineered features	92.0	0.929	0.894	0.911
Random Forest With engineered features	93.7	0.971	0.889	0.928
Logistic Regression	82.0	0.789	0.831	0.809

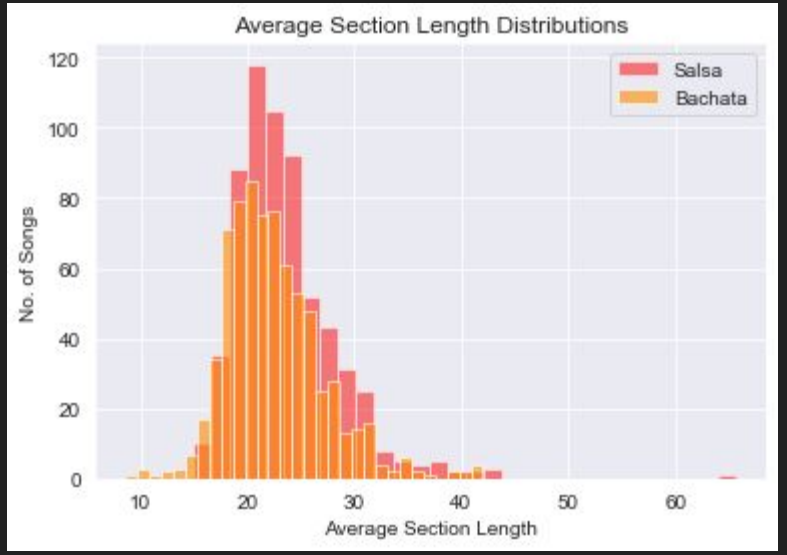
Hyperparameters tuned with GridSearchCV

What is wrong with the engineered features?

Don’t provide any distinguishing information.

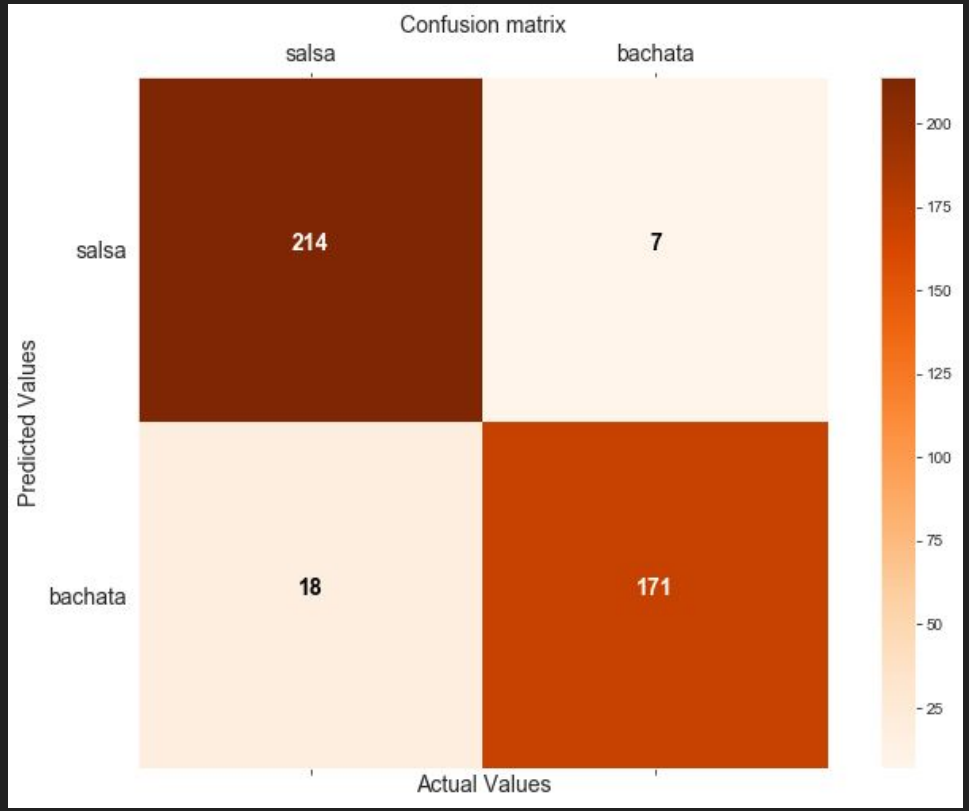
- Spotify’s algorithms need work



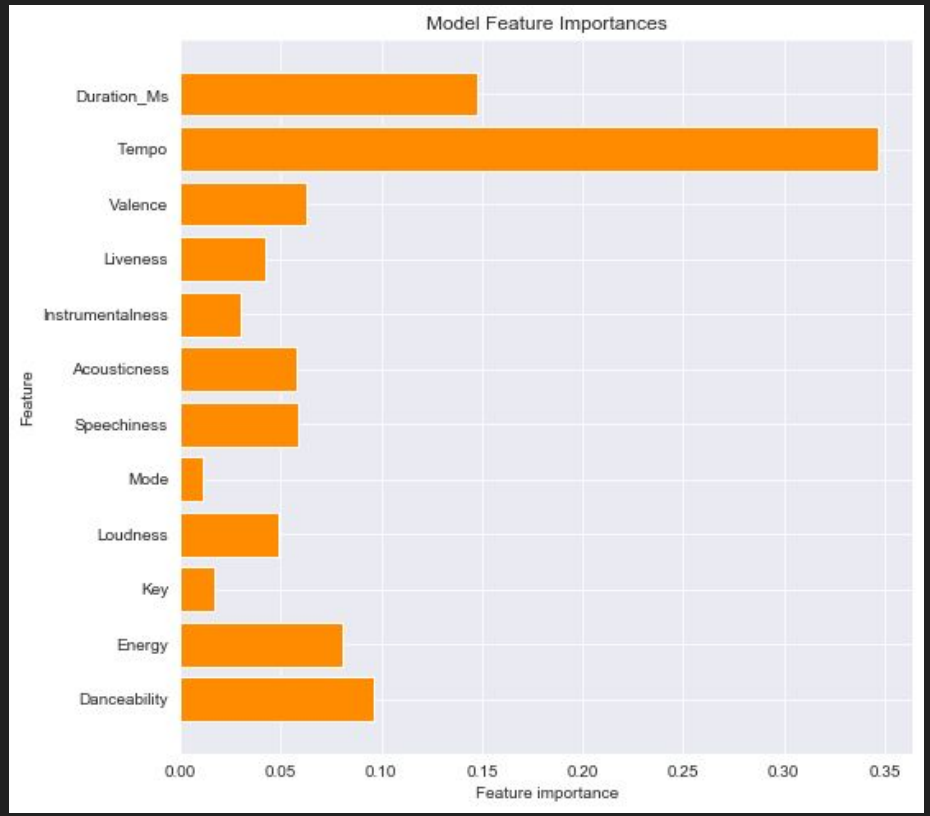


Preferred Model Performance

- **F1 = 0.932**
- Worse at characterizing salsa correctly
- Due to the fact that there are more salsa tracks in the test sample?



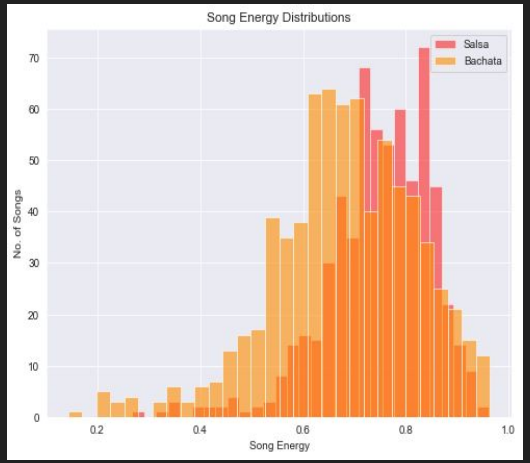
Feature Confusion

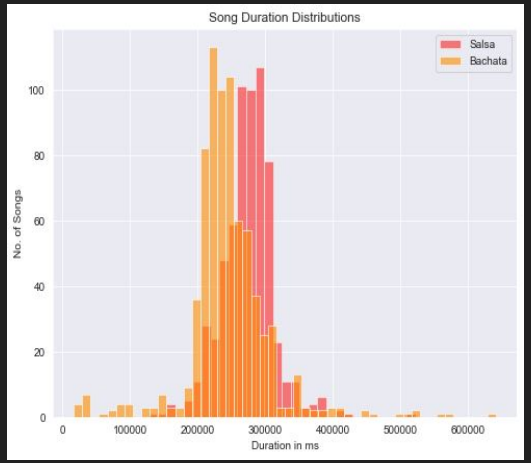


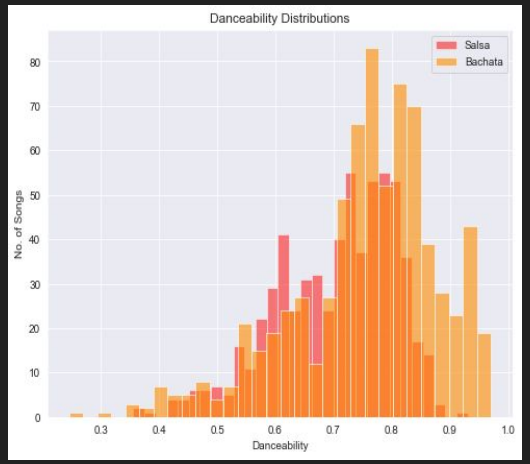
- Tempo
- Duration
- Danceability
- Energy

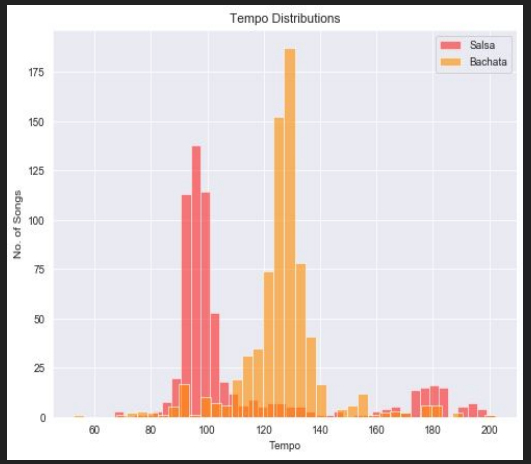
Feature Confusion

- Expect these incorrectly classified songs to lie within overlapping regions of most important model features



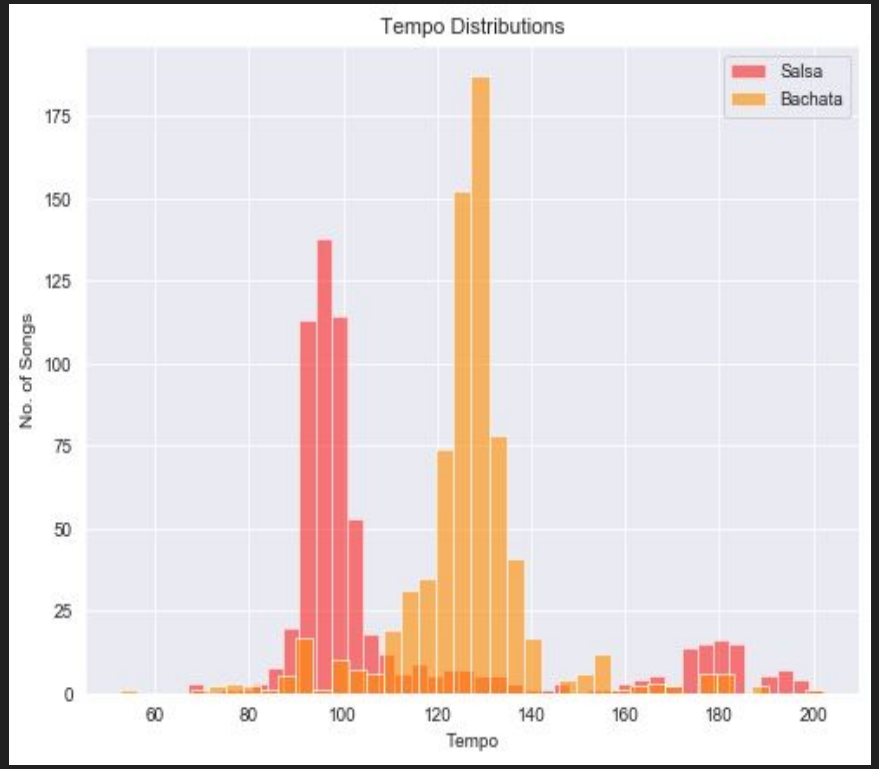






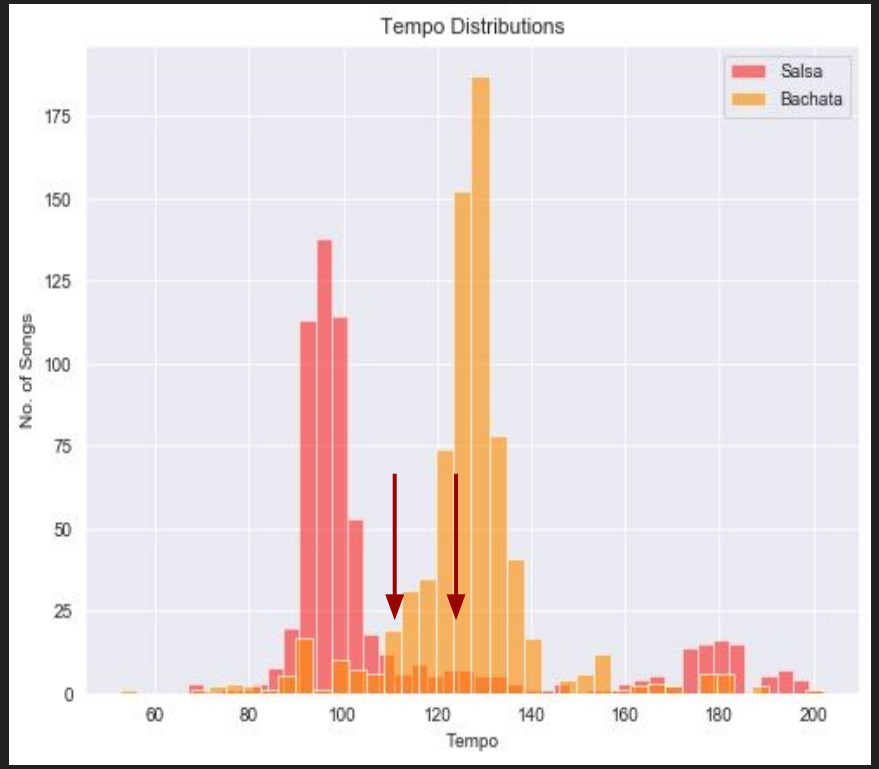
Avg. Incorrect Predictions vs. Avg. Statistics

genre	Full Dataset Tempo	Incorrect Tempo Predictions
bachata	126.399211	132.138500
salsa	111.137073	119.911765



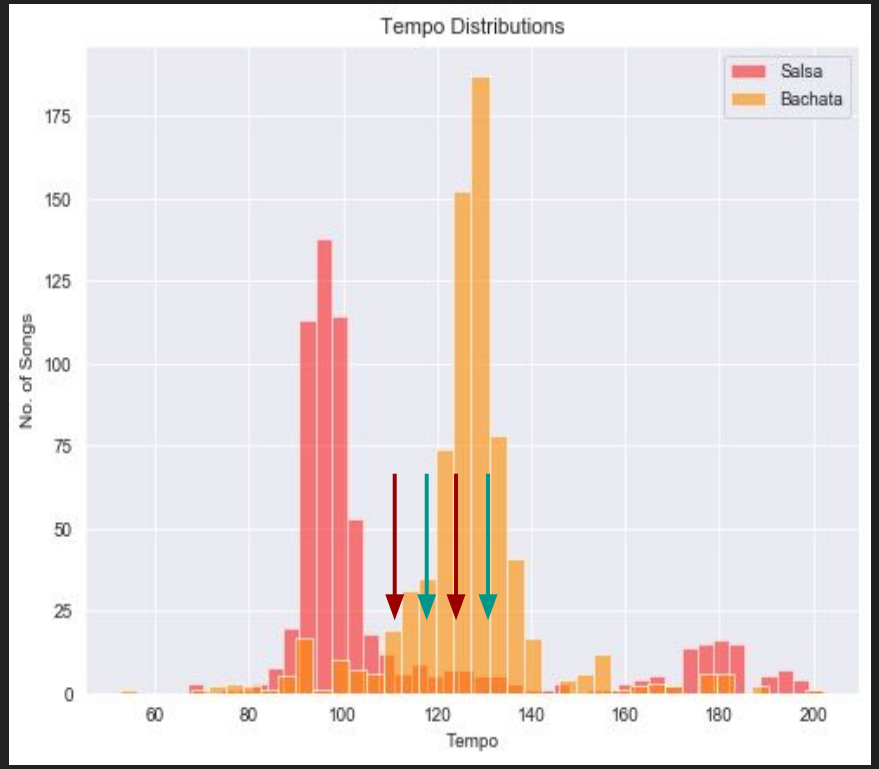
Avg. Incorrect Predictions vs. Avg. Statistics

genre	Full Dataset Tempo	Incorrect Tempo Predictions
bachata	126.399211	132.138500
salsa	111.137073	119.911765



Avg. Incorrect Predictions vs. Avg. Statistics

genre	Full Dataset Tempo	Incorrect Tempo Predictions
bachata	126.399211	132.138500
salsa	111.137073	119.911765



Avg. Incorrect Predictions vs. Avg. Statistics

genre	Full Dataset Tempo	Incorrect Tempo Predictions
bachata	126.399211	132.138500
salsa	111.137073	119.911765

genre	Full Dataset Duration	Incorrect Duration Predictions
bachata	244339.291156	231428.250000
salsa	275220.996820	257204.176471

genre	Full Dataset Energy	Incorrect Energy Predictions
bachata	0.678786	0.709000
salsa	0.747730	0.725765

genre	Full Dataset Danceability	Incorrect Danceability Predictions
bachata	0.755850	0.751625
salsa	0.702728	0.709529

Conclusion

- Salsa and bachata vary mainly in tempo, duration, danceability, and energy
- Model performs quite well: F1 = 0.932
  - However, incorrectly predicts bachata when the track is actually salsa more often than other way around
- Engineered features in this project were not useful
  - Spotify API “Audio Analysis” needs work
- More models

Appendix

Audio Features

Audio Analysis Section Object Features

PCA

Audio Features Object

KEY	VALUE TYPE	VALUE DESCRIPTION
acousticness	float	A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
analysis_url	string	An HTTP URL to access the full audio analysis of this track. An access token is required to access this data.
danceability	float	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.
duration_ms	int	The duration of the track in milliseconds.
energy	float	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.
id	string	The Spotify ID for the track.
instrumentalness	float	Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentalness value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.
key	int	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.
liveness	float	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.
loudness	float	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 typical range between -60 and 0 db.
mode	int	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	float	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.
tempo	float	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.
time_signature	int	An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure).
track_href	string	A link to the Web API endpoint providing full details of the track.
type	string	The object type: "audio_features"
uri	string	The Spotify URI for the track.
valence	float	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).

Section Object

KEY	VALUE TYPE	VALUE DESCRIPTION
start	float	The starting point (in seconds) of the section.
duration	float	The duration (in seconds) of the section.
confidence	float	The confidence, from `0.0` to `1.0`, of the reliability of the section's "designation".
loudness	float	The overall loudness of the section in decibels (dB). Loudness values are useful for comparing relative loudness of sections within tracks.
tempo	float	The overall estimated tempo of the section 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.
tempo_confidence	float	The confidence, from `0.0` to `1.0`, of the reliability of the *tempo*. Some tracks contain tempo changes or sounds which don't contain tempo (like pure speech) which would correspond to a low value in this field.
key	integer	The estimated overall key of the section. The values in this field ranging from `0` to `11` mapping 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`.
key_confidence	float	The confidence, from `0.0` to `1.0`, of the reliability of the *key*. Songs with many key changes may correspond to low values in this field.
mode	integer	Indicates the modality (major or minor) of a track, the type of scale from which its melodic content is derived. This field will contain a `0` for "minor", a `1` for "major", or a `-1` for no result. *Note that the major key (e.g. C major) could more likely be confused with the minor key at 3 semitones lower (e.g. A minor) as both keys carry the same pitches.*
mode_confidence	float	The confidence, from `0.0` to `1.0`, of the reliability of the *mode*.
time_signature	integer	An estimated overall time signature of a track. The time signature (meter) is a notational convention to specify how many beats are in each bar (or measure). The time signature ranges from `3` to `7` indicating time signatures of "3/4", to "7/4".
time_signature_confidence	float	The confidence, from `0.0` to `1.0`, of the reliability of the *time_signature*. Sections with time signature changes may correspond to low values in this field.

PCA

- Speechiness, acousticness, instrumentalness, and liveness all are correlated in the same direction
- Valence, tempo, mode, energy, and danceability are correlated in the same direction.
- Key and tempo don't really correlate much with those other variables in real life. They're pretty independent, so I think it's fair that their magnitudes are smaller.

	Loading
danceability	-0.269523
energy	-0.598041
key	-0.024909
loudness	-0.531406
mode	-0.114982
speechiness	0.473894
acousticness	0.212092
instrumentalness	0.035091
liveness	0.131718
valence	-0.242590
tempo	-0.030055
duration_ms	-0.134754

PCA

- Speechiness, acousticness, instrumentalness, and liveness all are correlated in the same direction
- Valence, tempo, mode, energy, and danceability are correlated in the same direction.
- Key and tempo don't really correlate much with those other variables in real life. They're pretty independent, so I think it's fair that their magnitudes are smaller.

	Loading
danceability	-0.269523
energy	-0.598041
key	-0.024909
loudness	-0.531406
mode	-0.114982
speechiness	0.473894
acousticness	0.212092
instrumentalness	0.035091
liveness	0.131718
valence	-0.242590
tempo	-0.030055
duration_ms	-0.134754