Spotify Data Analysis Report

Determining the Playlist Genre of a Song on Spotify Through Analysis

By

Eleanor Madderra, Sarayu Ramachandra, Sachi Satish Kumar, Snarr Pung, and Senait Pirani

December 5, 2024

Table of Contents

Introduction

We are analyzing a dataset ('spotify_songs.csv' that has been downloaded from Kaggle) that contains information about 28,356 songs on Spotify that are in playlists that are reported to be 5 different genres (pop, r&b, latin, rap, and edm).

We want to be able to predict the genre of a playlist a song is in based on data about the song including danceability, energy, key, loudness, speechiness, acousticness, instrumentalness, valence, and tempo. These variables are described below.

Our Data

Variables

- **Danceability**
    - Double
    - Quantitative variable
    - 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.
    - Some genres might be considered more 'danceable' than others. We may be able to predict the genre of the playlist a song is in depending on how 'danceable' a track is considered to be.
- **Energy**
    - Double
    - Quantitative variable
    - 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.
    - Some genres might be considered more energetic than others. We may be able to predict the genre of the playlist a song is in depending on how energetic a track is considered to be.
- **Key**
    - Double
    - Quantitative variable

    - The estimated overall key of the track. 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.
    - Some genres might use one key more than other genres. We may be able to predict the genre of the playlist a song is in depending on what key a song is in.
- **Loudness**
    - Double
    - Quantitative variable
    - 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.
    - Some genres might be louder than others. We may be able to predict the genre of the playlist a song is in depending on how loud a track is.
- **Speechiness**
    - Double
    - Quantitative variable
    - 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.
    - Some genres might contain more spoken words than others. We may be able to predict the genre of the playlist a song is in depending on the ‘speechiness’ of a track.
- **Acousticness**
    - Double
    - Quantitative variable
    - A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.
    - Some genres might be more likely to be acoustic than others. We may be able to predict the genre of the playlist a song is in depending on the acoustic confidence of a track.
- **Instrumentalness**
    - Double
    - Quantitative variable
    - 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.
    - Some genres might contain more instrumentals than others. We may be able to predict the genre of the playlist a song is in depending on the 'instrumentalness' of a track.
- **Valence**
    - Double
    - Quantitative variable
    - 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).
    - Some genres might be considered more positive than others. We may be able to predict the genre of the playlist a song is in depending on the valence of a track.
- **Tempo**
    - Double
    - Quantitative variable
    - 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.
    - Some genres could tend to have a faster tempo than others. We may be able to predict the genre of the playlist a song is in depending on the tempo of a track.
- **Playlist_genre**
    - Character
    - Categorical variable
    - The genre of the playlist of the song
    - This is how we will be able to determine genre.

Genre Stats

Below is our data sorted by genre. The tables display the count, mean, standard deviation, minimum value, 25%, 50%, 75%, and maximum value of each variable within the specified genre.

POP	dance-ability	energy	key	loud-ness	mode	speech-iness	acousticness	instru-mental-ness	valence	tempo
count	5507.00	5507.00	5507.00	5507.00	5507.00	5507.00	5507.00	5507.00	5507.00	5507.00
mean	0.63930	0.70102	5.31886	-6.3153	0.58816	0.07399	0.17079	0.05987	0.50352	120.743
std	0.12822	0.17108	3.63919	2.62220	0.49221	0.06782	0.21935	0.18364	0.22047	24.7484
min	0.09850	0.00814	0.00000	-26.279	0.00000	0.02280	0.00000	0.00000	0.02760	35.4770
25%	0.56300	0.59400	2.00000	-7.5080	0.00000	0.03690	0.01750	0.00000	0.33900	102.987
50%	0.65200	0.72700	5.00000	-5.8350	1.00000	0.04900	0.07670	0.00001	0.50000	120.017
75%	0.72900	0.83000	9.00000	-4.5690	1.00000	0.07910	0.23100	0.00213	0.66800	130.089
max	0.97900	0.99900	11.0000	-0.7000	1.00000	0.86900	0.99200	0.98200	0.98100	212.137

R&B	dance-ability	energy	key	loud-ness	mode	speech-iness	acousticness	instru-mental-ness	valence	tempo
count	5431.000000	5431.000000	5431.000000	5431.000000	5431.000000	5431.000000	5431.000000	5431.000000	5431.000000	5431.000000
mean	0.670179	0.590934	5.400479	-7.864848	0.521451	0.116792	0.259904	0.028920	0.531231	114.222156

std	0.138213	0.179407	3.596447	2.891189	0.499586	0.107182	0.256324	0.121776	0.225883	28.638985
min	0.140000	0.011800	0.000000	-34.283000	0.000000	0.022400	0.000025	0.000000	0.036600	46.169000
25%	0.584000	0.469000	2.000000	-9.397000	0.000000	0.041900	0.049000	0.000000	0.353500	92.977000
50%	0.689000	0.596000	6.000000	-7.420000	1.000000	0.067900	0.165000	0.000005	0.542000	108.744000
75%	0.771000	0.721000	9.000000	-5.832000	1.000000	0.158000	0.412500	0.000586	0.709000	129.931500
max	0.977000	0.995000	11.000000	-0.478000	1.000000	0.918000	0.989000	0.969000	0.990000	214.516000

Latin	**dance-ability**	**energy**	**key**	**loud-ness**	**mode**	**speech-iness**	**acousticness**	**instru-mental-ness**	**valence**	**tempo**
count	5155.000000	5155.000000	5155.000000	5155.000000	5155.000000	5155.000000	5155.000000	5155.000000	5155.000000	5155.000000
mean	0.713287	0.708312	5.483996	-6.264455	0.561979	0.102653	0.210920	0.044447	0.605510	118.622354
std	0.114974	0.152308	3.636293	2.865150	0.496192	0.087729	0.213611	0.168453	0.222289	29.130579

min	0.077100	0.000175	0.000000	-46.448000	0.000000	0.023200	0.000036	0.000000	0.000010	48.718000
25%	0.655000	0.620000	2.000000	-7.406500	0.000000	0.043600	0.044500	0.000000	0.443500	95.989000
50%	0.729000	0.729000	6.000000	-5.752000	1.000000	0.067400	0.139000	0.000002	0.628000	110.962000
75%	0.792500	0.821000	9.000000	-4.435500	1.000000	0.127000	0.309000	0.000329	0.784500	129.004000
max	0.979000	1.000000	11.000000	-0.046000	1.000000	0.662000	0.989000	0.994000	0.976000	239.440000

Rap	**dance-ability**	**energy**	**key**	**loud-ness**	**mode**	**speech-iness**	**acousticness**	**instru-mental-ness**	**valence**	**tempo**
count	5746.000000	5746.000000	5746.000000	5746.000000	5746.000000	5746.000000	5746.000000	5746.000000	5746.000000	5746.000000
mean	0.718353	0.650708	5.470936	-7.042269	0.521406	0.197506	0.192479	0.075997	0.505090	120.654908
std	0.136452	0.170340	3.697661	3.054460	0.499585	0.132444	0.218727	0.229977	0.224663	31.644504
min	0.150000	0.016100	0.000000	-26.207000	0.000000	0.024300	0.000002	0.000000	0.029200	38.985000

25%	0.634000	0.546000	2.000000	-8.391500	0.000000	0.078825	0.030000	0.000000	0.333000	92.773250
50%	0.737000	0.665000	6.000000	-6.507500	1.000000	0.177500	0.109000	0.000000	0.517000	119.987500
75%	0.819750	0.776000	9.000000	-5.047250	1.000000	0.290000	0.281000	0.000182	0.680000	145.041000
max	0.975000	0.999000	11.000000	0.642000	1.000000	0.877000	0.994000	0.974000	0.977000	211.644000

EDM	**dance-ability**	**energy**	**key**	**loud-ness**	**mode**	**speech-iness**	**acousticness**	**instru-mental-ness**	**valence**	**tempo**
count	6043.000000	6043.000000	6043.000000	6043.000000	6043.000000	6043.000000	6043.000000	6043.000000	6043.000000	6043.000000
mean	0.655041	0.802476	5.352143	-5.427445	0.520106	0.086695	0.081504	0.218578	0.400656	125.768024
std	0.123558	0.139415	3.557832	2.371431	0.499637	0.071058	0.145439	0.326492	0.226205	15.321600
min	0.162000	0.106000	0.000000	-19.563000	0.000000	0.023900	0.000003	0.000000	0.026900	60.045000
25%	0.576000	0.723000	2.000000	-6.554000	0.000000	0.043800	0.003770	0.000006	0.217000	123.011500

How to Run Notebook

Ensure that 'spotify_songs.csv' is in the same directory as 'final_notebook_group_4_v2.ipynb' and not a sub directory.

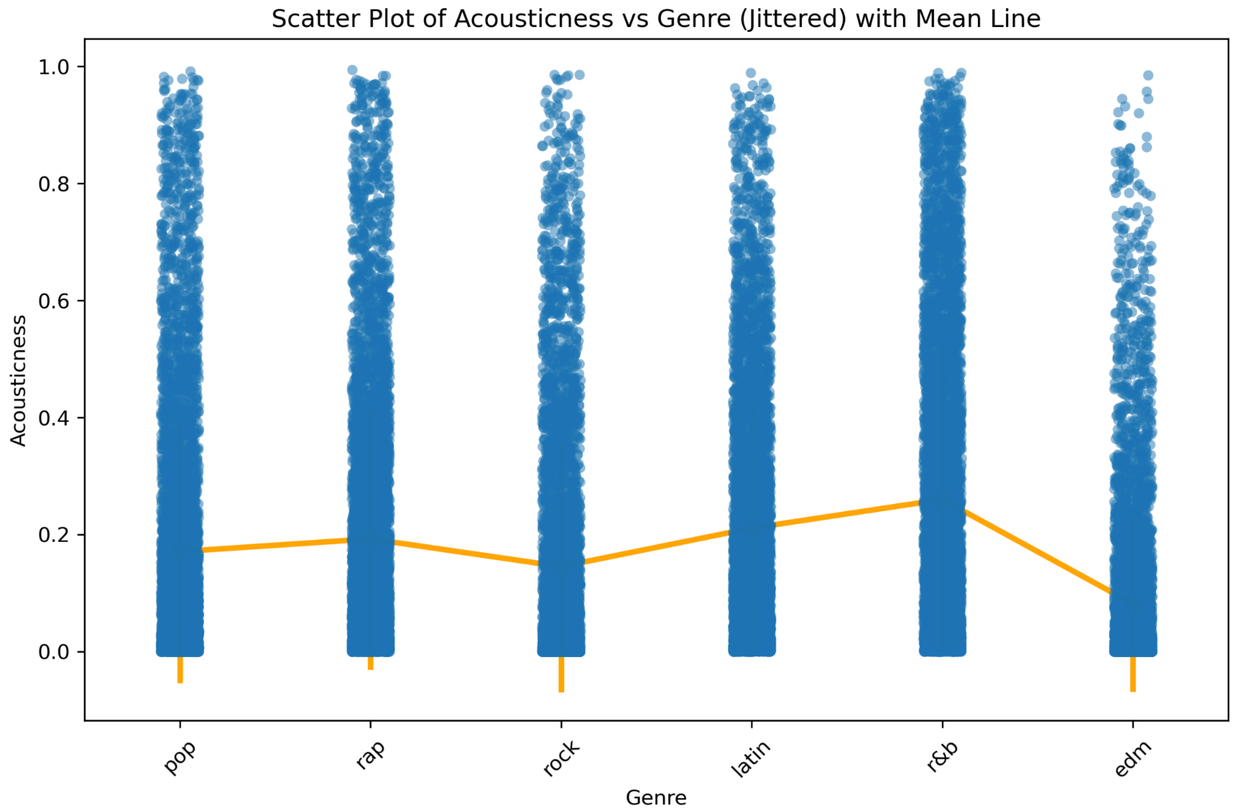
Note: There may be times where you need to restart the kernel to use updated packages.

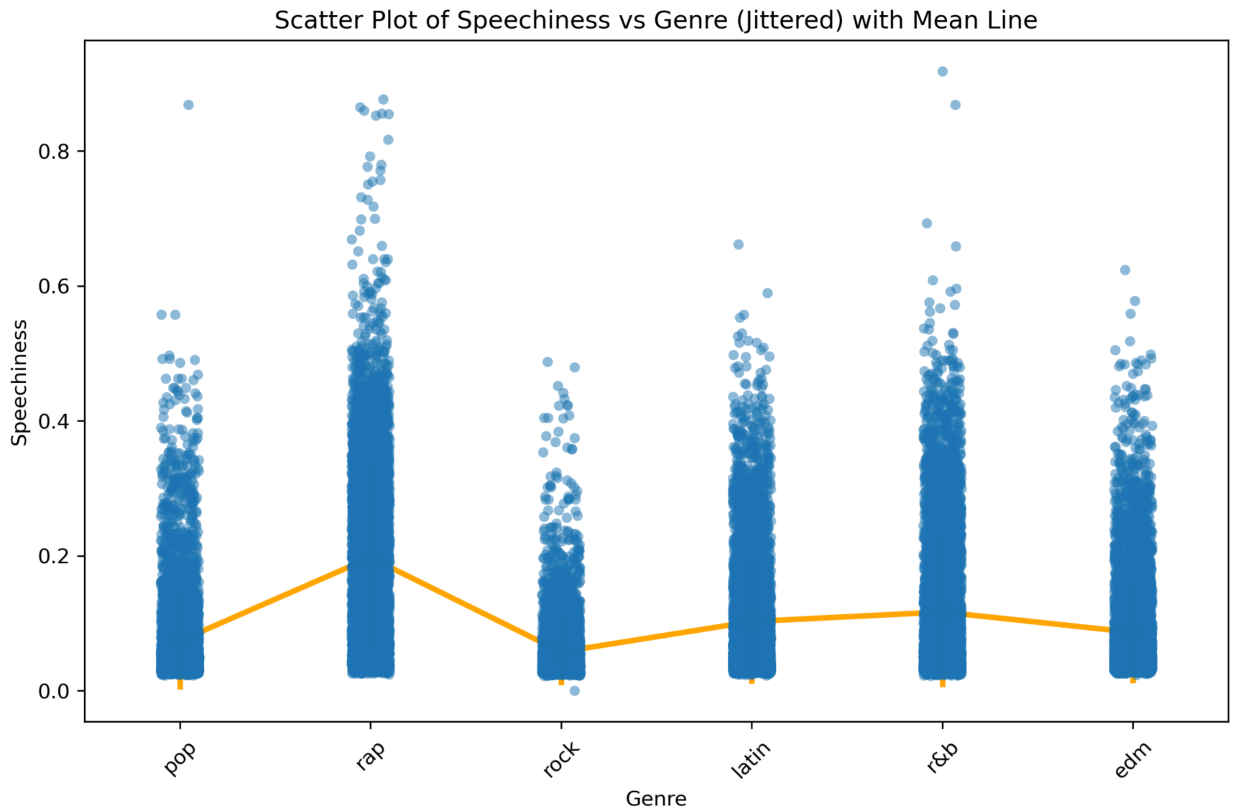
Note: The SVM model may take a while to fully load.

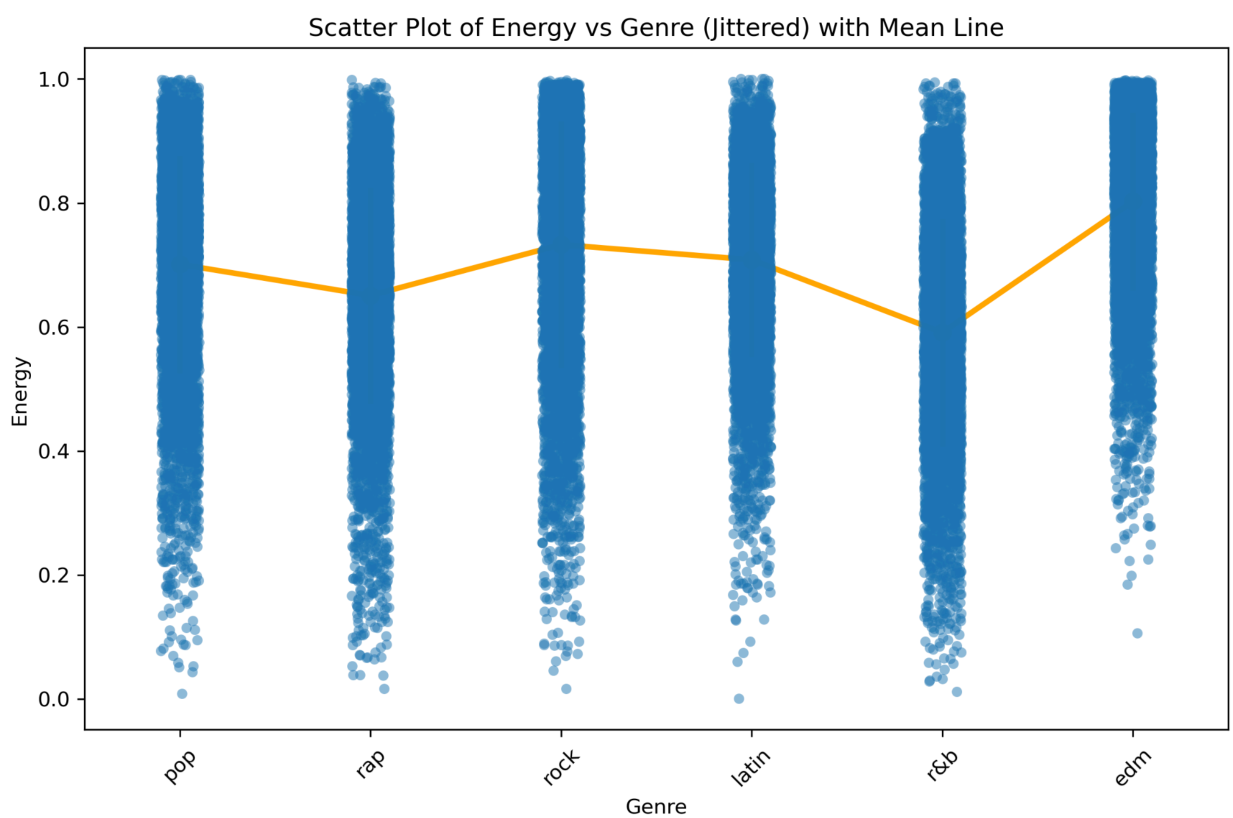
Our Figures

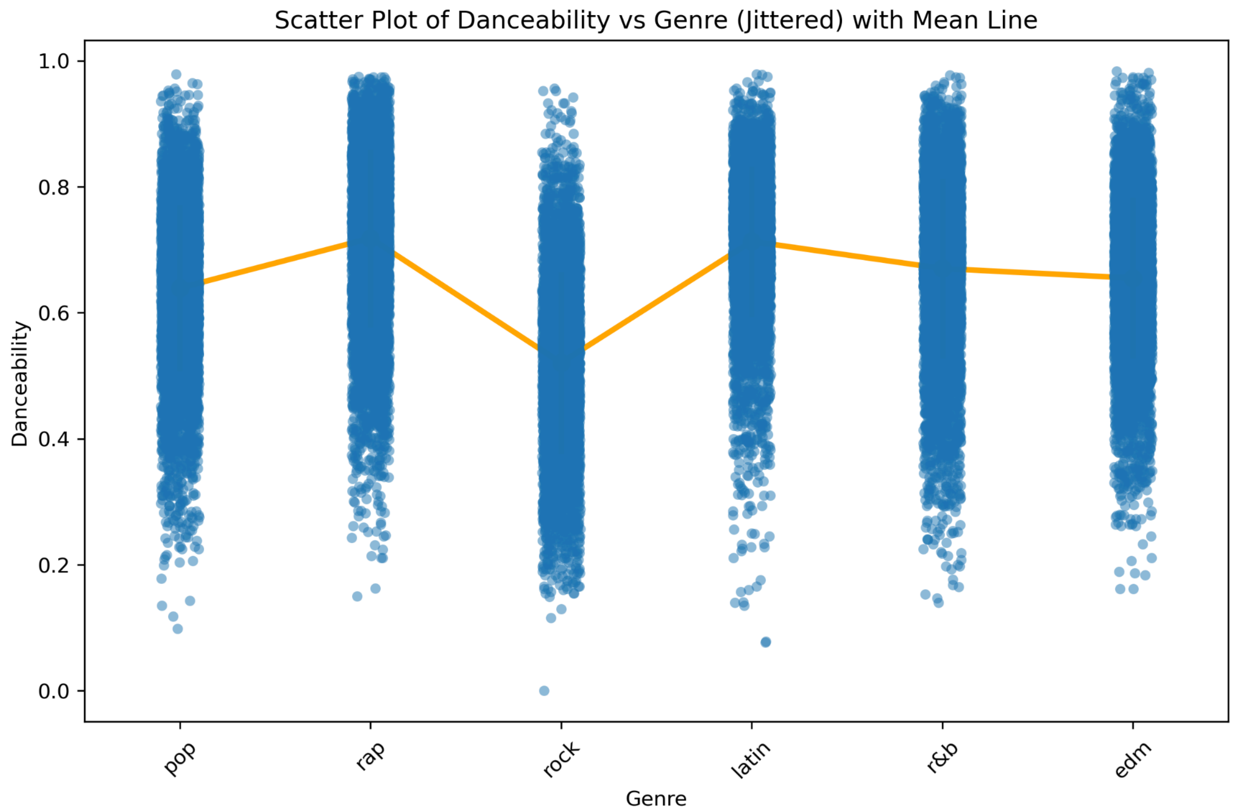
Five figure types were analyzed: scatter plots, histograms, bar charts, pie charts, and a radar chart. These figures were used to get an overall feel of the data set and to determine which variables may have an effect on the playlist genre and which variables may not. Variables that do not have an effect on playlist genre will be excluded from the training of our models in an effort to be more efficient.

Scatter Plot

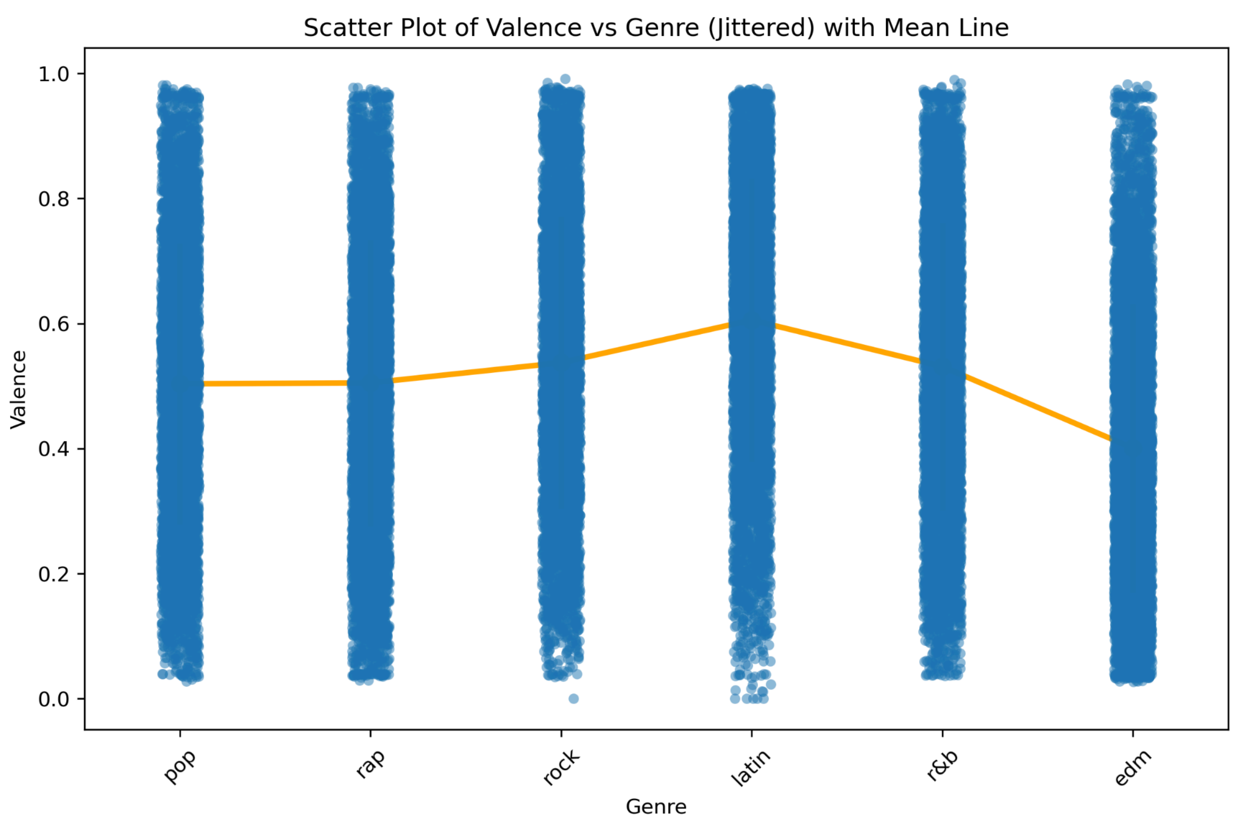



Scatter Plot of Speechiness vs Genre (Jittered) with Mean Line
Speechiness
0.0
0.2
0.4
0.6
0.8
pop
rap
rock
latin
r&b
edm
Genre


Scatter Plot of Energy vs Genre (Jittered) with Mean Line
Energy
0.0
0.2
0.4
0.6
0.8
1.0
pop
rap
rock
latin
r&b
edm
Genre



Scatter Plot of Danceability vs Genre (Jittered) with Mean Line
Danceability
Genre
pop
rap
rock
latin
r&b
edm
0.0
0.2
0.4
0.6
0.8
1.0

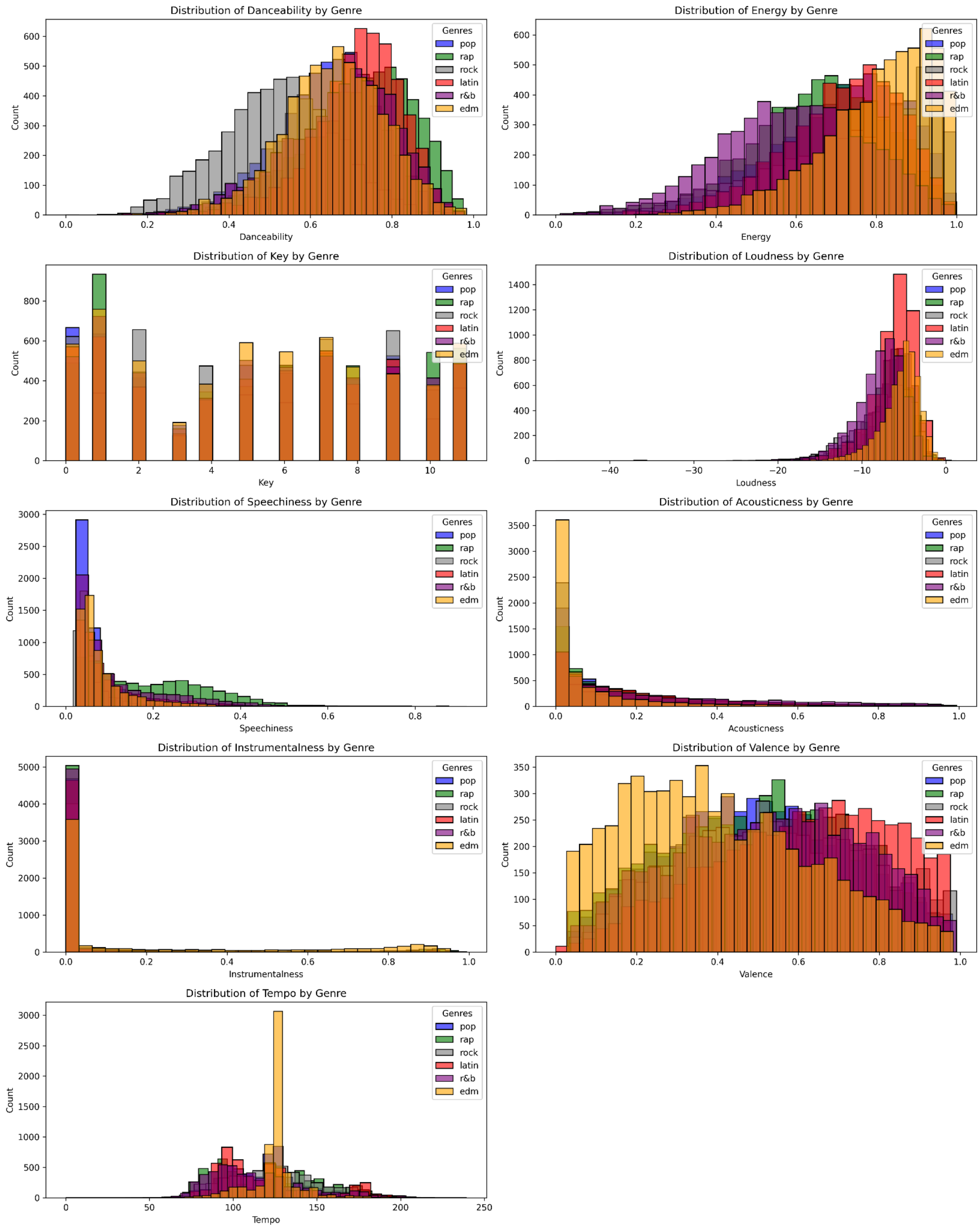


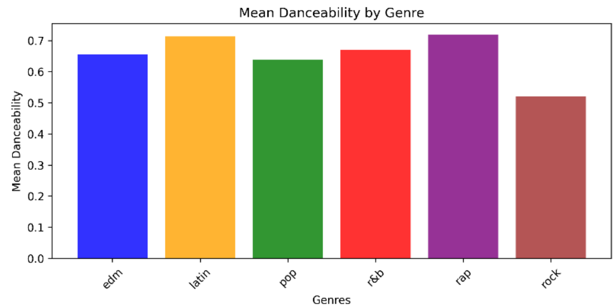
Scatter Plot of Valence vs Genre (Jittered) with Mean Line
Valence
Genre
pop
rap
rock
latin
r&b
edm
0.0
0.2
0.4
0.6
0.8
1.0

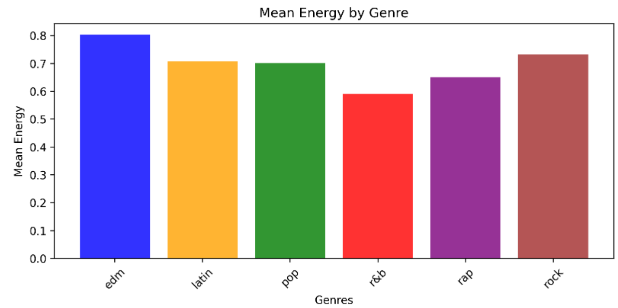
Interpretation

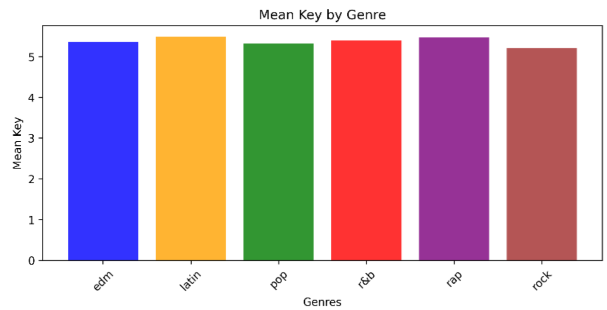
The above jittered scatterplots with mean trend lines show how our chosen variables vary across different genres. Valence shows that pop and Latin music tend to have more upbeat tracks, while genres such as rap and EDM have lower valence averages. Speechiness is distinctly higher in rap, showing the genre's focus on spoken content, while pop and rock show lower levels. Energy trends highlight EDM and Latin as more energetic genres, while rock and R&B appear to be less energetic on average. Pop, Latin, and EDM rank higher in danceability, while rock and rap are slightly lower. Acousticness is more pronounced in R&B and rock, while EDM shows the least acousticness, which is consistent with its electronic production method.

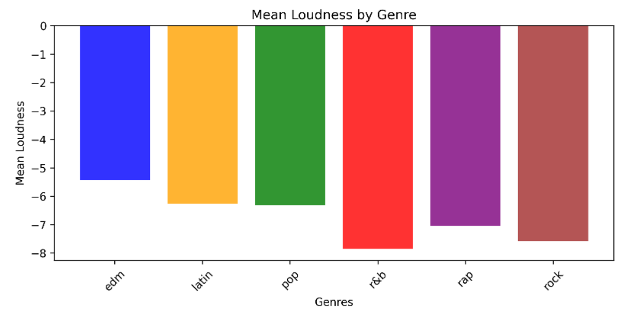
Histogram

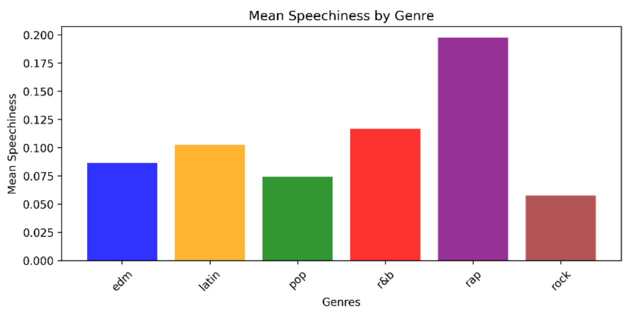

Distribution of Danceability by Genre
Danceability
Count
Distribution of Energy by Genre
Energy
Distribution of Key by Genre
Key
Distribution of Loudness by Genre
Loudness
Distribution of Speechiness by Genre
Speechiness
Distribution of Acousticness by Genre
Acousticness
Distribution of Instrumentalness by Genre
Instrumentalness
Distribution of Valence by Genre
Valence
Distribution of Tempo by Genre
Tempo
Genres
pop
rap
rock
latin
r&b
edm

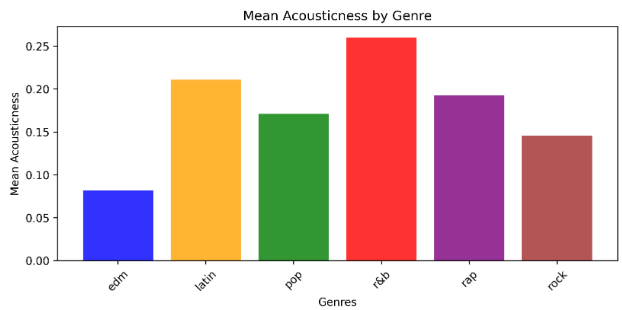

Mean Danceability by Genre
Mean Danceability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
edm
latin
pop
r&b
rap
rock
Genres

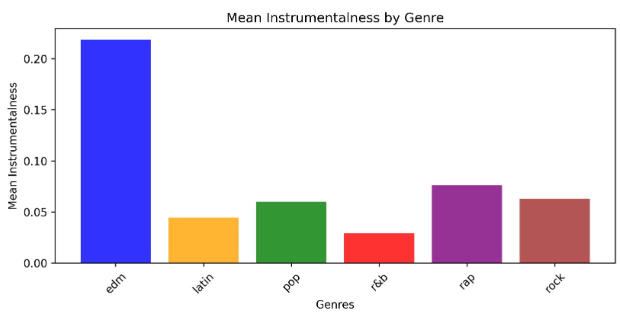

Mean Energy by Genre
Mean Energy
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
edm
latin
pop
r&b
rap
rock
Genres

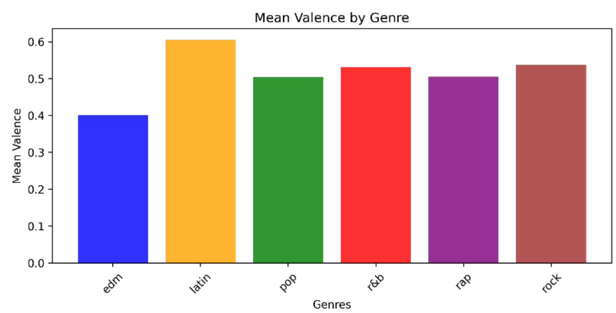

Mean Key by Genre
Mean Key
0
1
2
3
4
5
edm
latin
pop
r&b
rap
rock
Genres

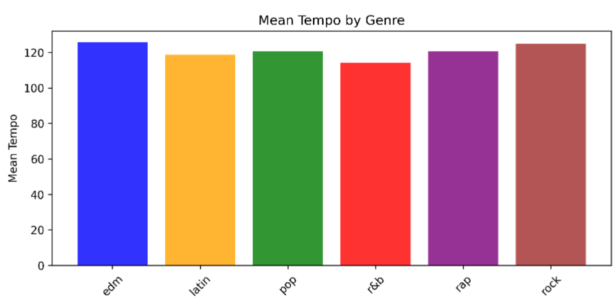

Mean Loudness by Genre
Mean Loudness
0
−1
−2
−3
−4
−5
−6
−7
−8
edm
latin
pop
r&b
rap
rock
Genres


Mean Speechiness by Genre
Mean Speechiness
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
edm
latin
pop
r&b
rap
rock
Genres


Mean Acousticness by Genre
Mean Acousticness
0.00
0.05
0.10
0.15
0.20
0.25
edm
latin
pop
r&b
rap
rock
Genres


Mean Instrumentalness by Genre
Mean Instrumentalness
0.00
0.05
0.10
0.15
0.20
edm
latin
pop
r&b
rap
rock
Genres


Mean Valence by Genre
Mean Valence
0.0
0.1
0.2
0.3
0.4
0.5
0.6
edm
latin
pop
r&b
rap
rock
Genres

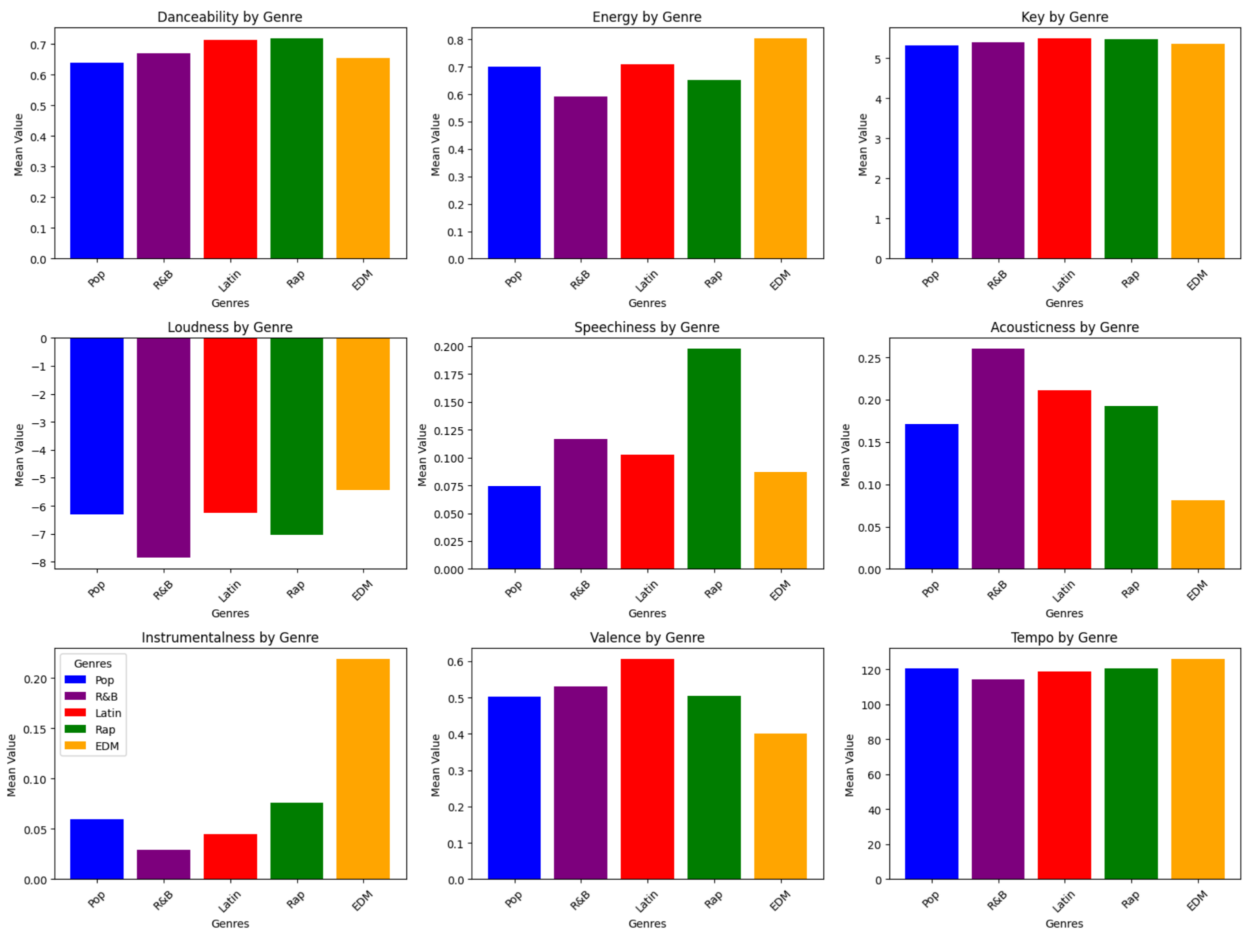

Mean Tempo by Genre
Mean Tempo
0
20
40
60
80
100
120
edm
latin
pop
r&b
rap
rock
Genres

Interpretation

Histograms are a powerful tool for visualizing data distributions and are crucial for analyzing the Spotify dataset in this project. By dividing data into bins and counting the frequency of values within each bin, histograms allow us to understand the spread and concentration of features such as danceability, energy, tempo, and speechiness across different genres. This analysis reveals key patterns and trends that help differentiate genres based on their unique characteristics. For example, EDM and pop genres exhibit high mean values for danceability, energy, and loudness, while rap stands out for its higher speechiness, which reflects its focus on lyrical content. Additionally, genres like R&B and Latin show higher acousticness, highlighting their use of natural instruments compared to the more electronic-driven sounds of EDM and rap.

The histograms also provide valuable insights into feature relevance and distribution. Features like energy, speechiness, and tempo show clear genre-specific patterns, making them strong candidates for predicting genre based on song attributes. For instance, EDM has a high mean tempo and energy, while rap tends to have slower tempos and lower energy. Meanwhile, key distributions are relatively uniform, suggesting that key may not be as significant for genre prediction. Overall, these histograms not only help us understand the general behavior of each feature within a genre but also reveal the critical variables that can be used for genre classification. By examining these patterns, we gain deeper insights into the relationships between song features and genre characteristics, laying the foundation for building effective predictive models.

Bar Chart



Interpretation

These bar charts are helpful in our research into this dataset. These charts are displaying the means of each of the features and allows us to compare the means between genres. Going through each feature individually, we can see that danceability has mean values that are relatively close between genres, however the Latin and Rap genres have slightly higher values. Energy has a bit more variation, with EDM being the highest and R&B being the lowest, with a difference of approximately .22. The keys are all relatively the same, and has very little variation throughout the genres. In loudness, we can see more variation. We see a great amount of variation between genres in the speechiness feature. We can see that Rap has the highest speechiness, which is logical considering how rap songs are faster paced and are characterized by a strong flow of words. In Acousticness, we can see the greatest high under R&B, and the lowest value under EDM. Instrumentalness has the greatest difference between the highest and lowest value, with EDM being the highest and R&B being the lowest. Valence has less variation, and Tempo has even less. Based on these values, we can conclude that features like instrumentalness, acousticness, and speechiness have the most variation and will be the most helpful in determining and distinguishing between genres like EDM, Rap, and R&B

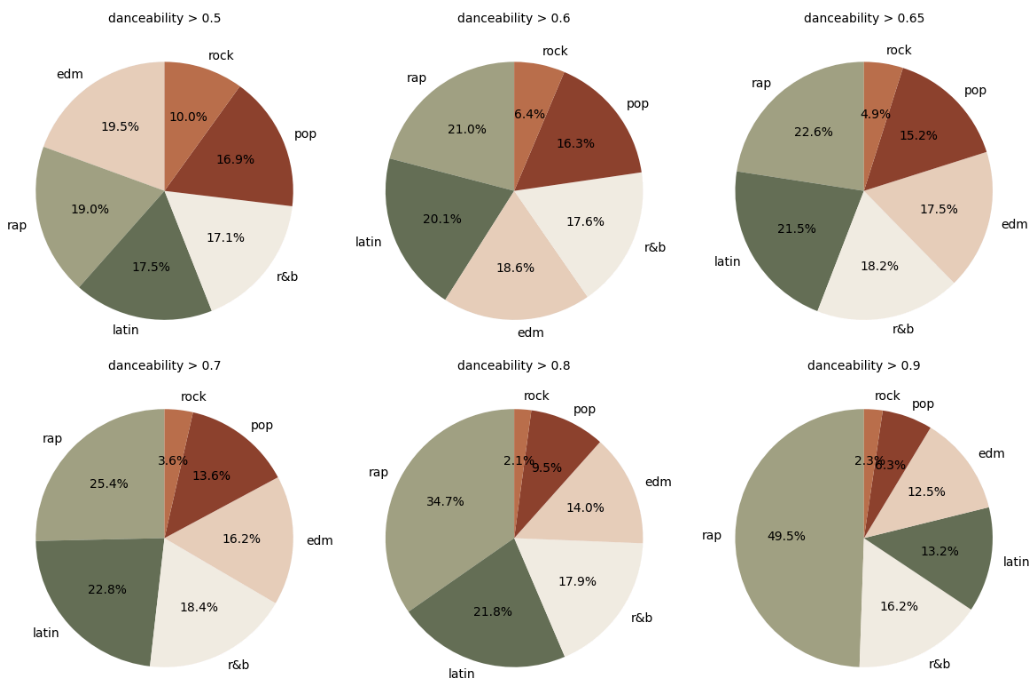
Pie Chart

Multiple pie charts were created for each variable, as shown below. Depending on the variable, 6 to 12 pie charts were created to represent the data when the variable is above a specified threshold (the value displayed at the top of the chart) and another 6 to 12 charts were made for when the variable falls below a specified threshold. The thresholds analyzed were determined based on the range of means of a specified variable and the overall range. These pie charts illustrate the distribution of playlist genres when the variable is above or below a specified threshold.

Danceability

Range Mean: 0.639 - 0.718

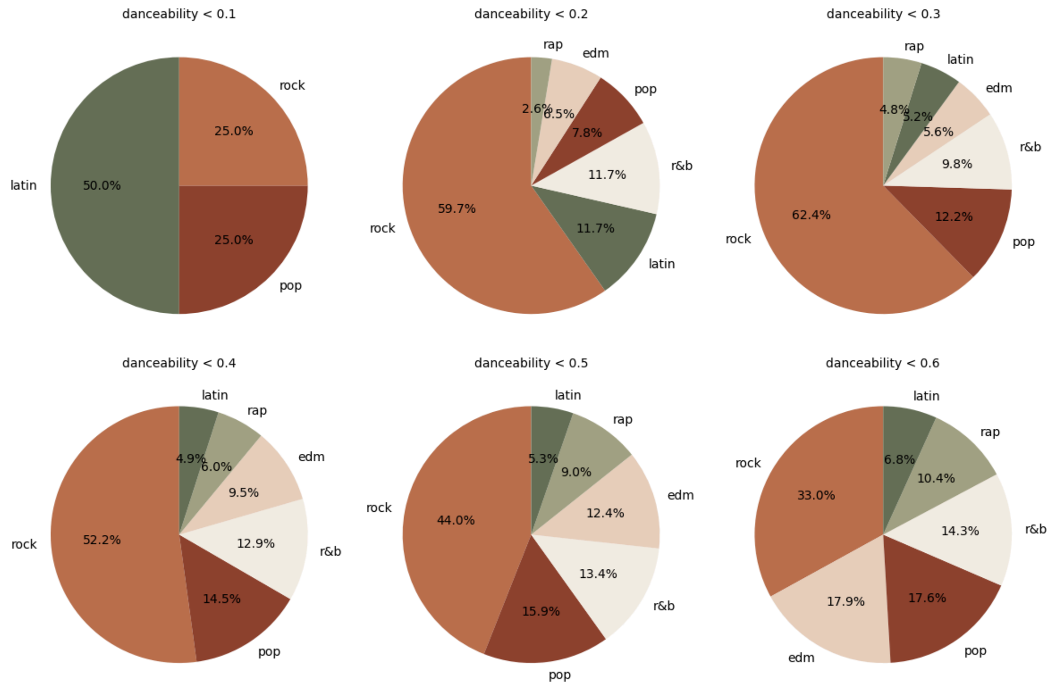
Range: 0.07 - 0.983



As seen in the pie charts above, rock is unlikely to occur when danceability is above 0.6 and continues to occur less and less as the danceability variable increases.

When danceability is above 0.8, pop occurs less.

When danceability is above 0.8, rap occurs more often.



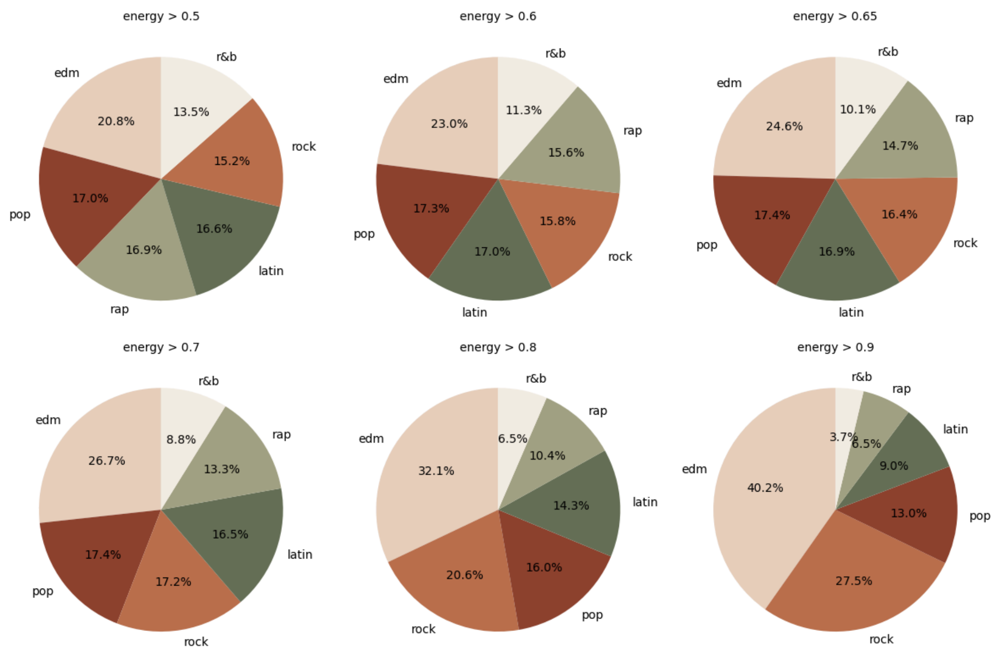
When danceability is below 0.1, there are no occurrences of edm, r&b, or rap.

When danceability is below 0.4, rock is most likely.

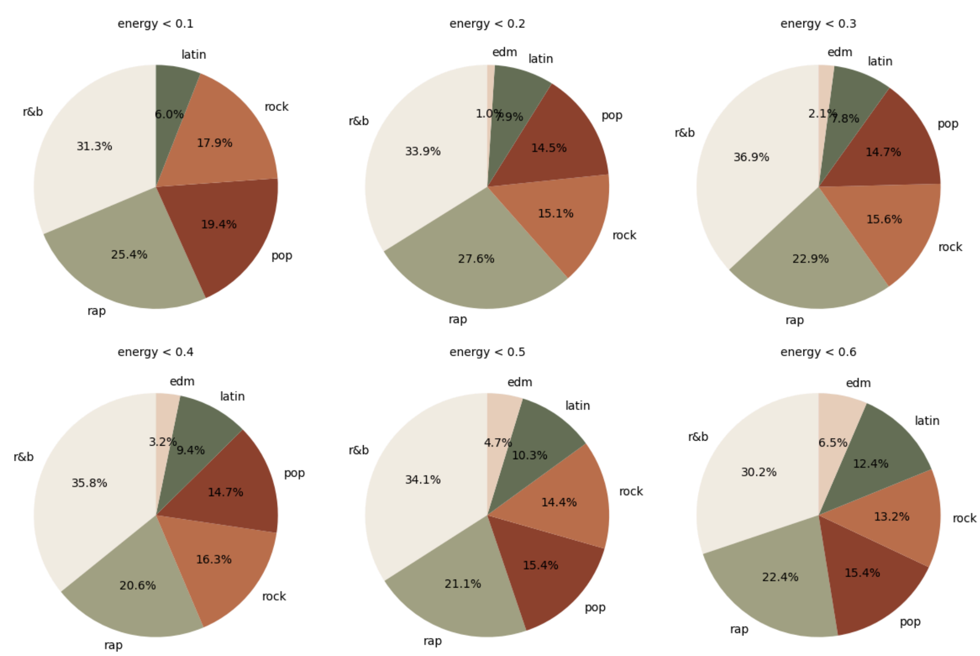
Energy

Range Mean: 0.590 - 0.802

Range: 0.0001 - 1



When energy is greater than 0.9, edm is most likely.

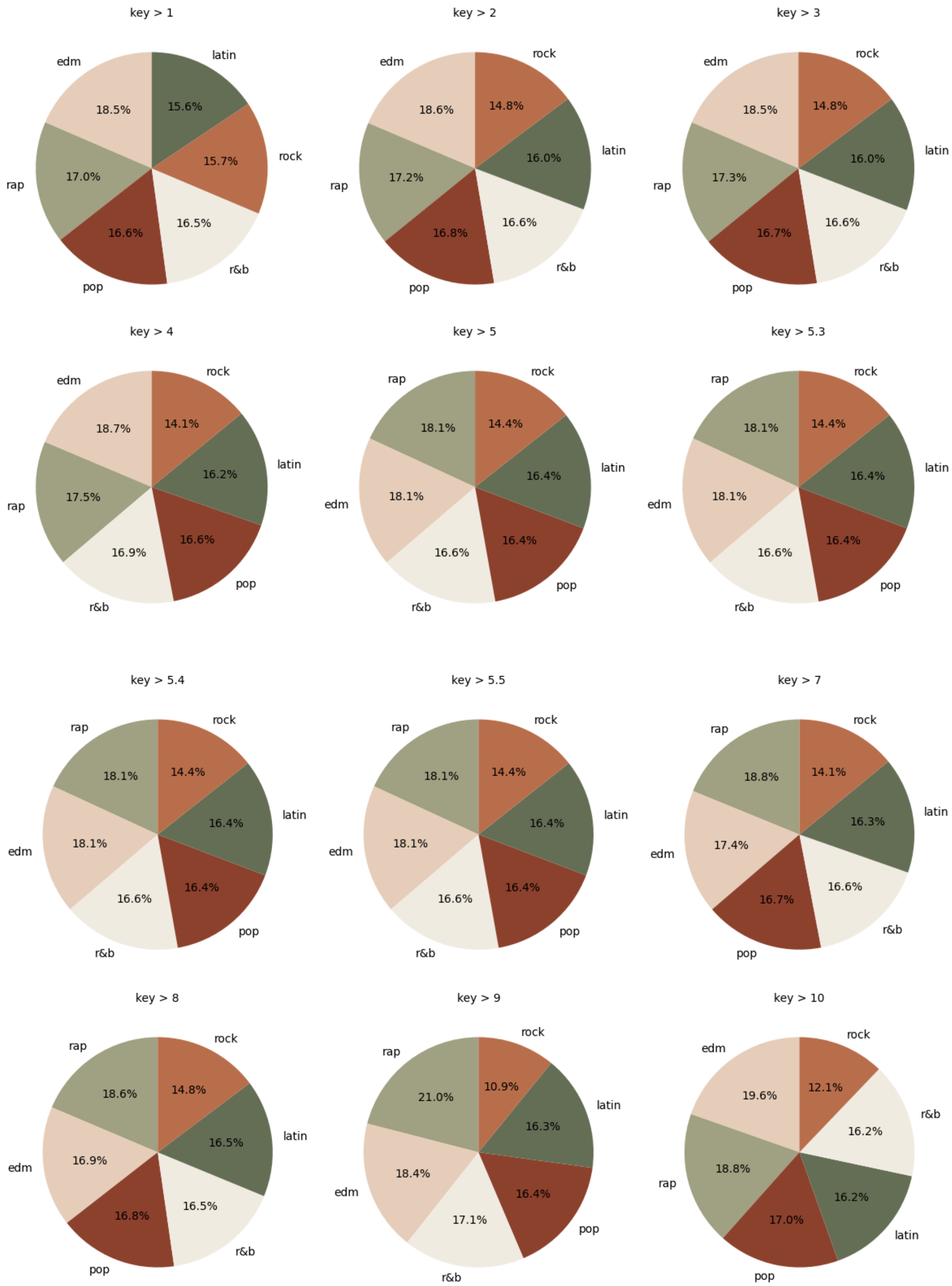


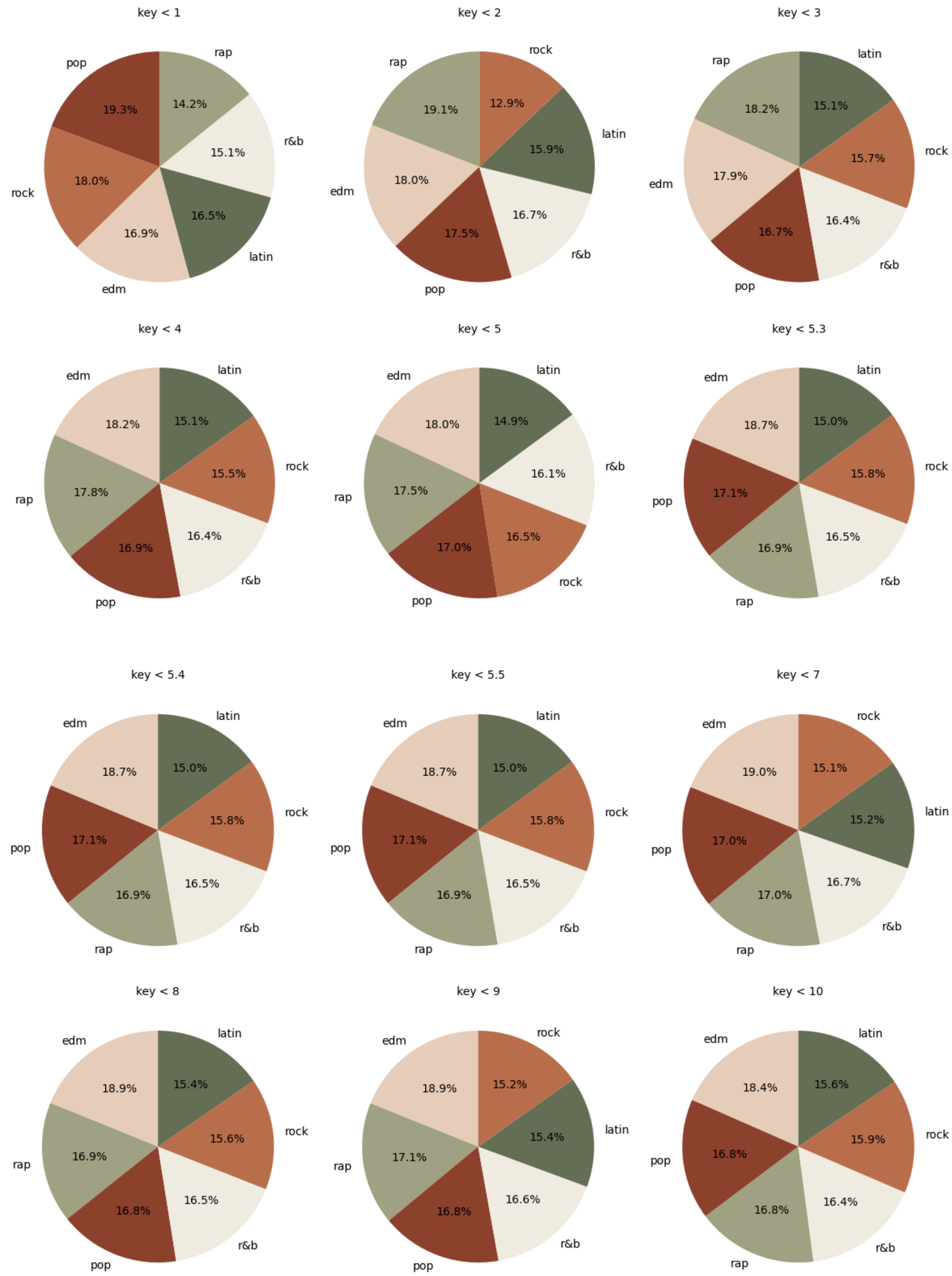
When energy is less than 0.6, R&B is likely and edm is not.

Key

Range Mean: 5.31 - 5.48

Range: 0 - 11



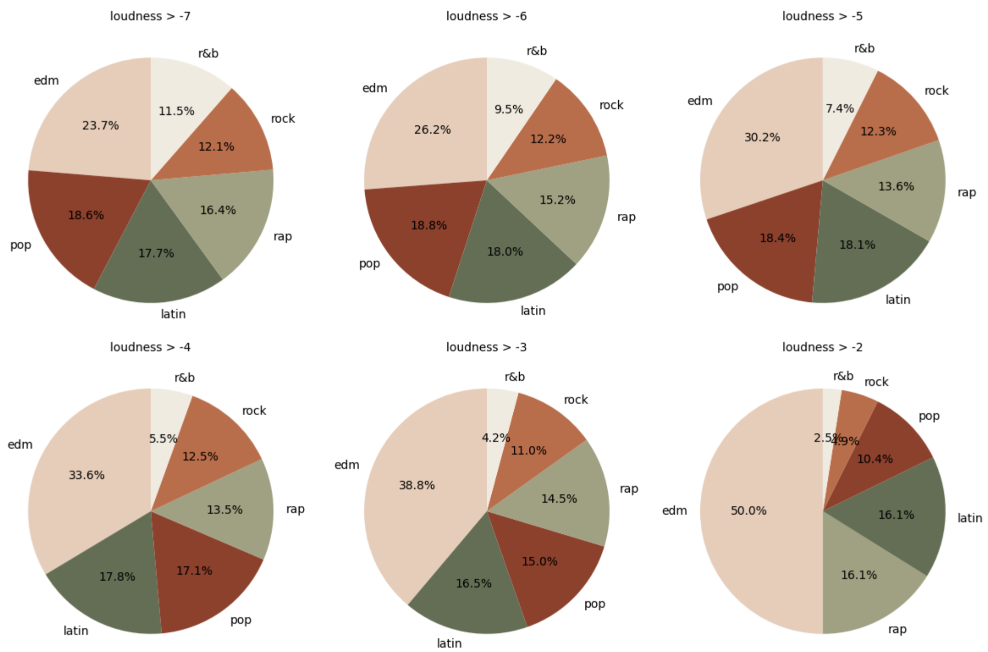


Key does not seem to affect the likeliness of a genre occurring or not. The distribution of genres is pretty evenly split.

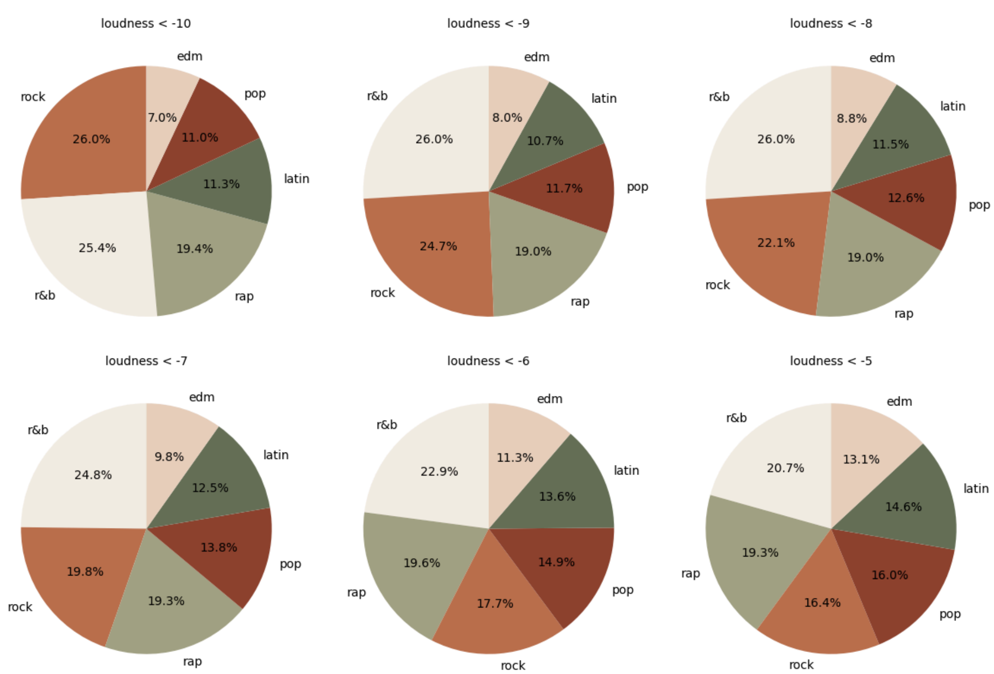
Loudness

Range Mean: -5.427 - -7.864

Range: -46.4 - 1.135



When loudness is greater than -2 but less than -4, edm is most likely.

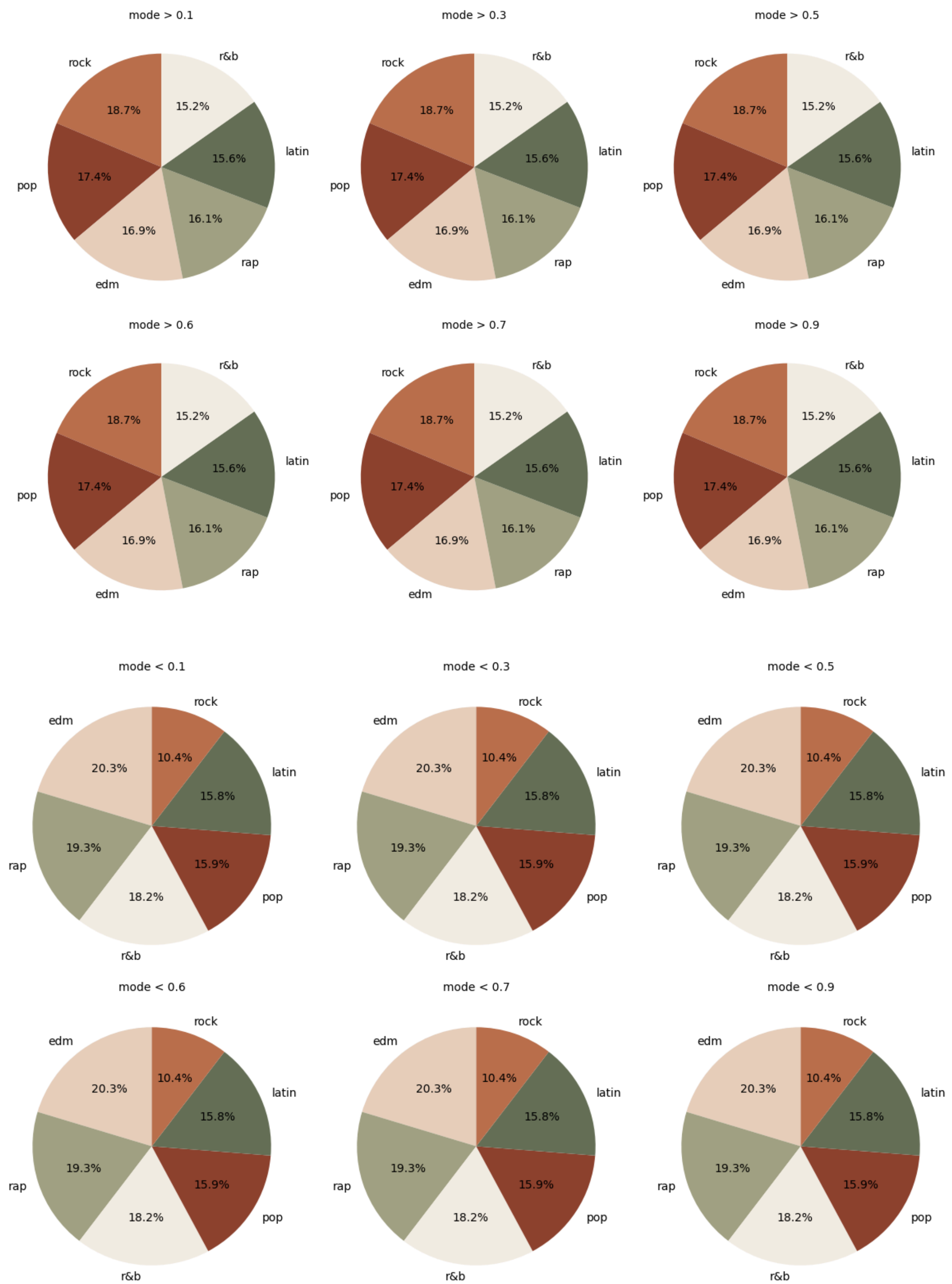


When loudness is greater than -8, edm is not as likely.

Mode

Range Mean: 0.5201 - 0.588

Range: 0 - 1

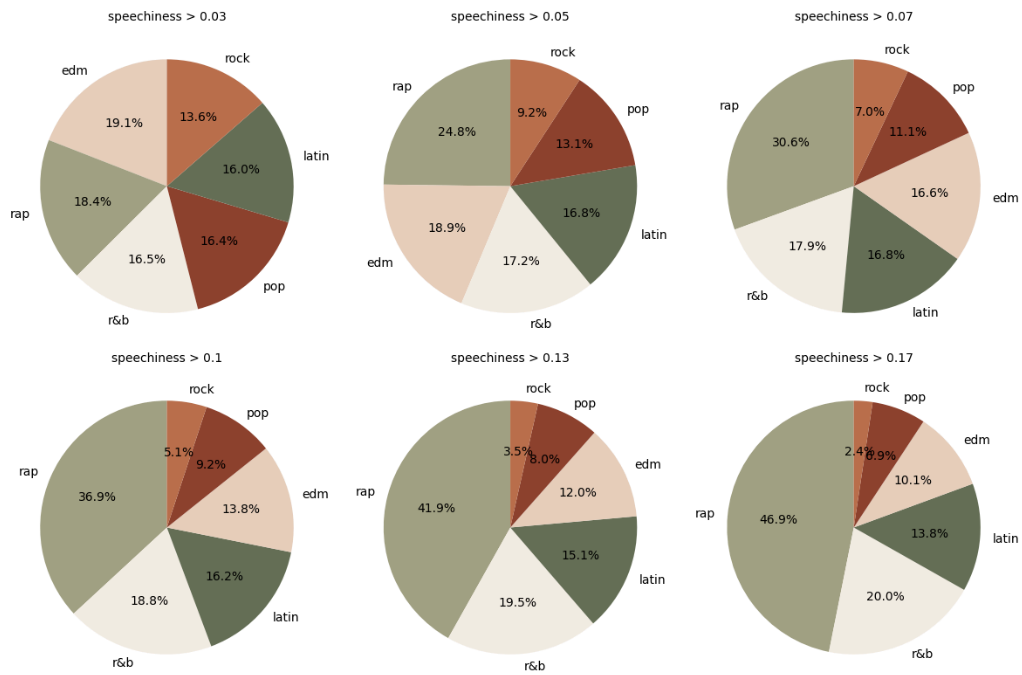


Mode does not seem to have a great effect on the likeliness of a genre occurring or not. The distribution of genres is pretty evenly split.

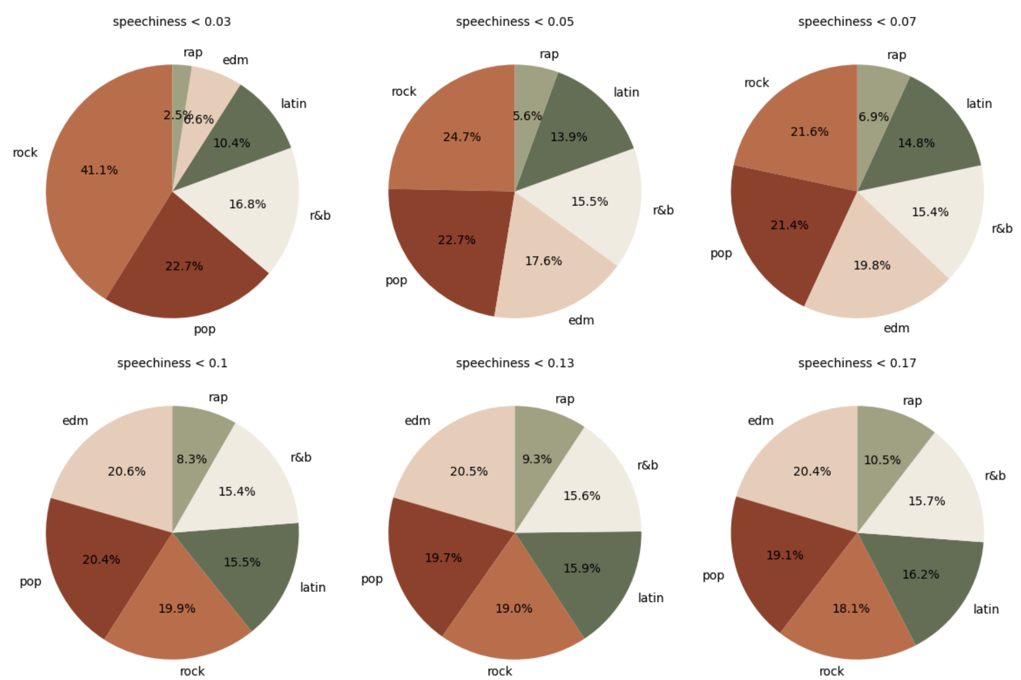
Speechiness

Range Mean: 0.0739 - 0.1975

Range: 0.02 - 0.91



When speechinees is over 0.07, rock and pop are not as likely.

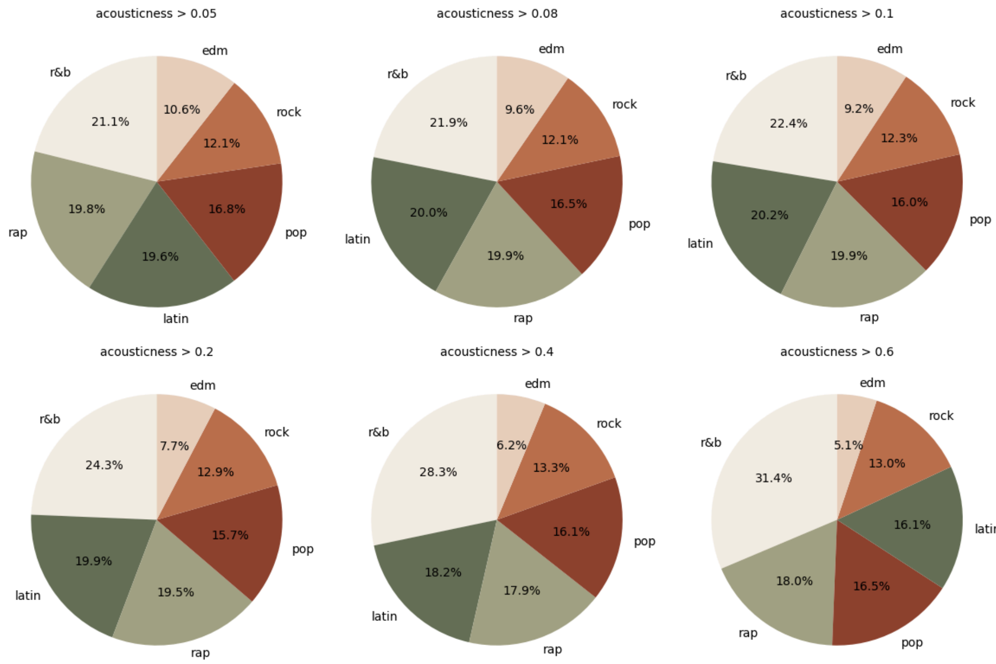


When speechiness is less than 0.03, rap, edm, and latin are not as likely.

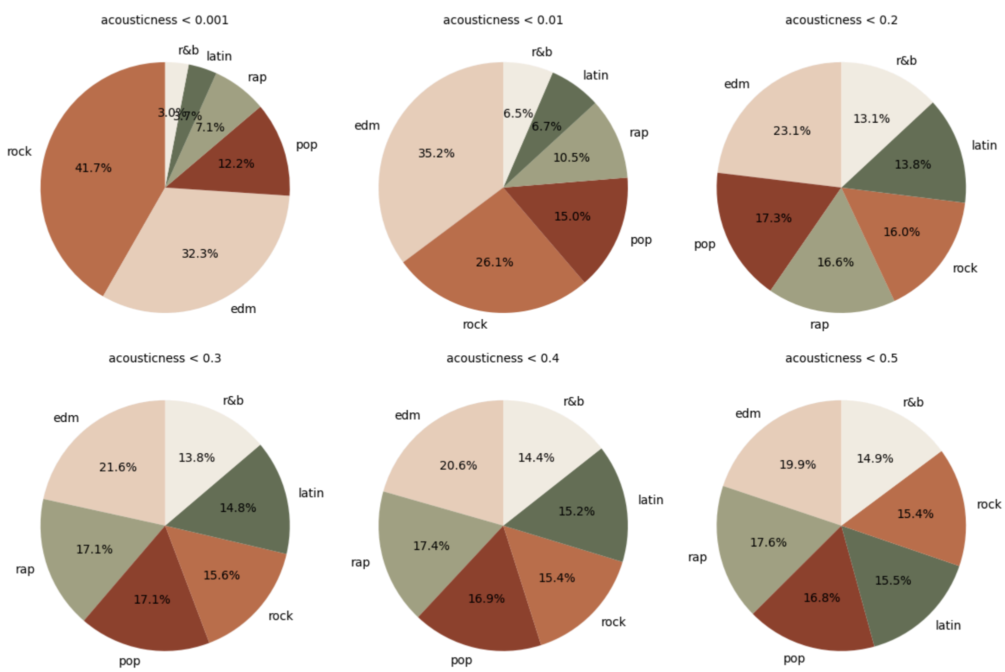
Acousticness

Range Mean: 0.0815 - 0.2599

Range: 0.000002 - 0.989



When acousticness is greater than 0.6, edm is less likely.

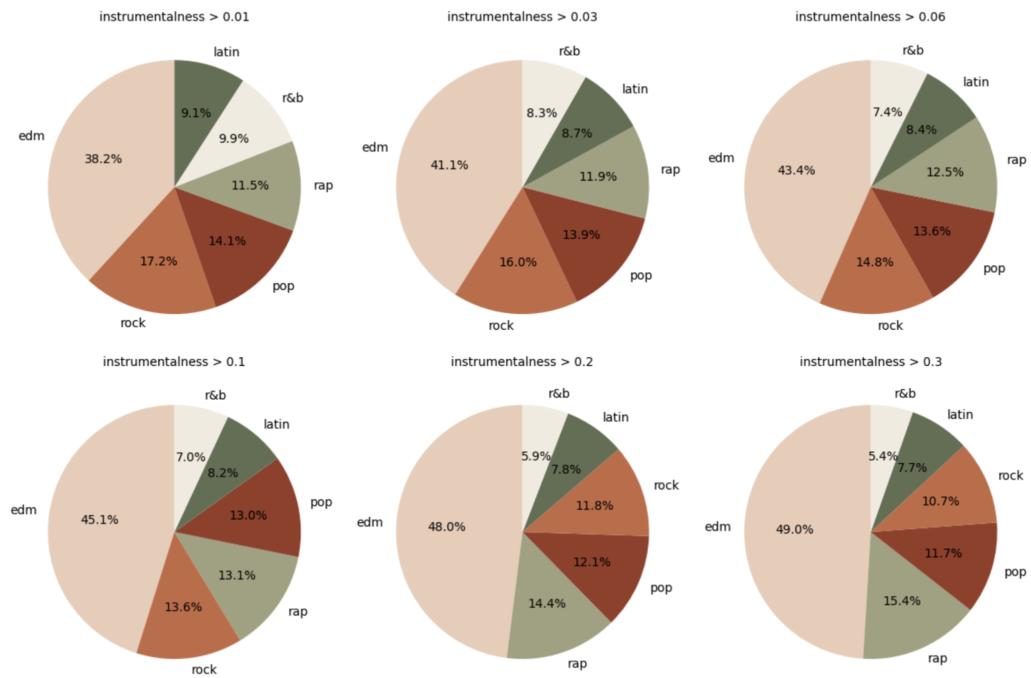


When acousticness is less than 0.0001, rock and edm are most likely.

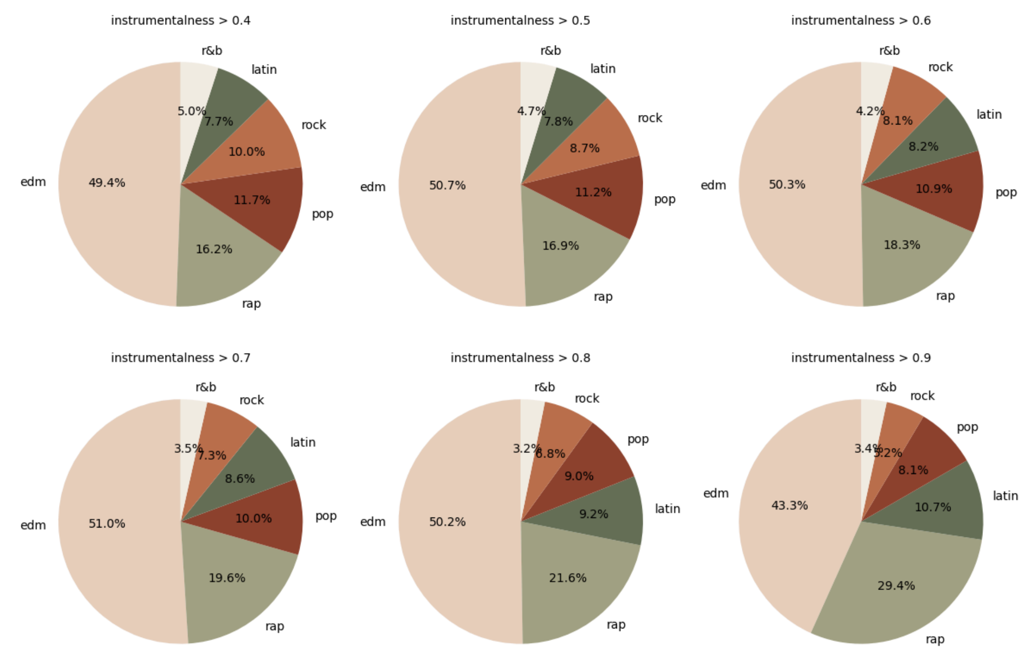
Instrumentalness

Range Mean: 0.0289 - 0.2185

Range: 0 - 0.984



When instrumentalness is greater than 0.1, edm is most likely.

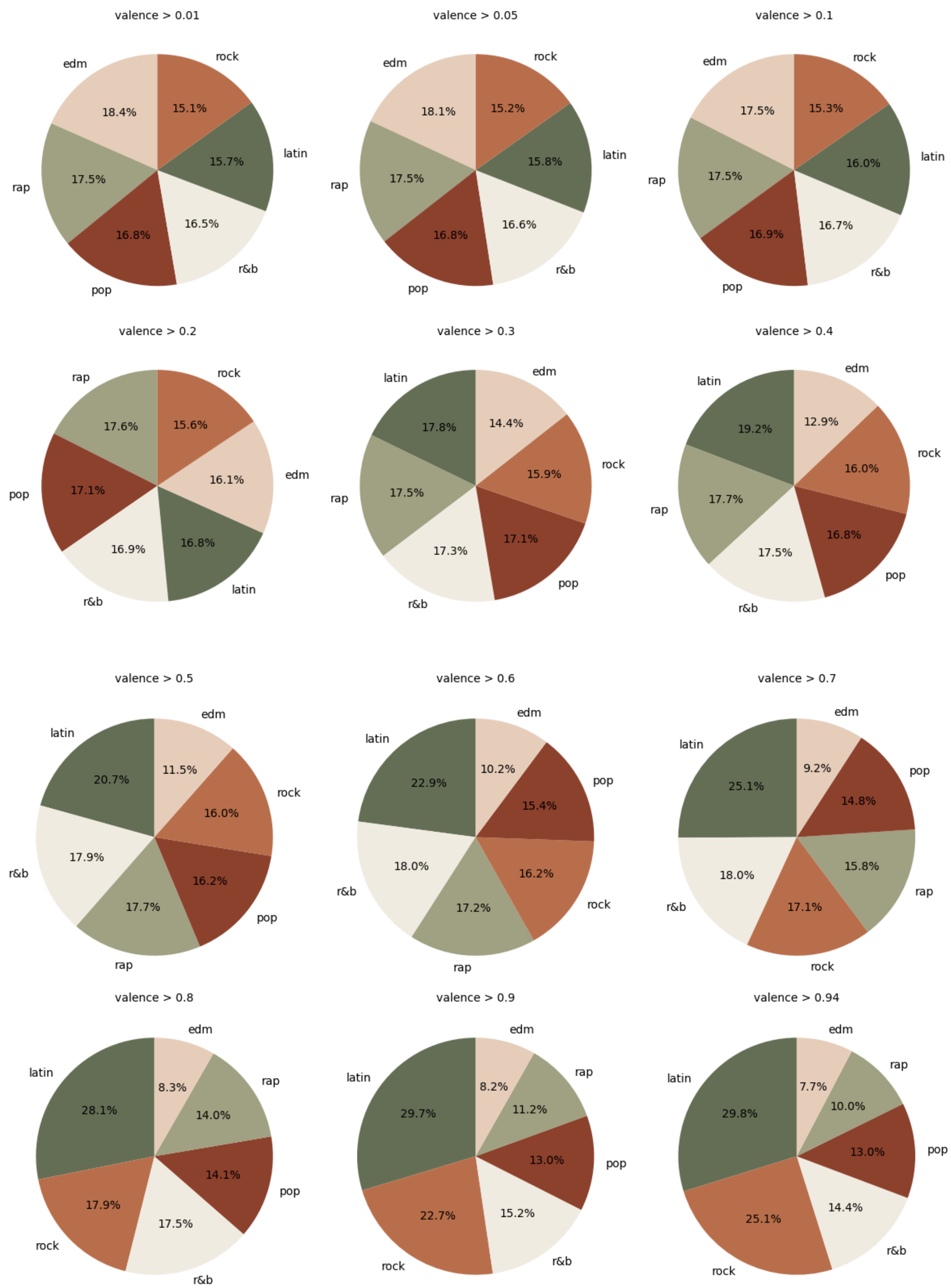


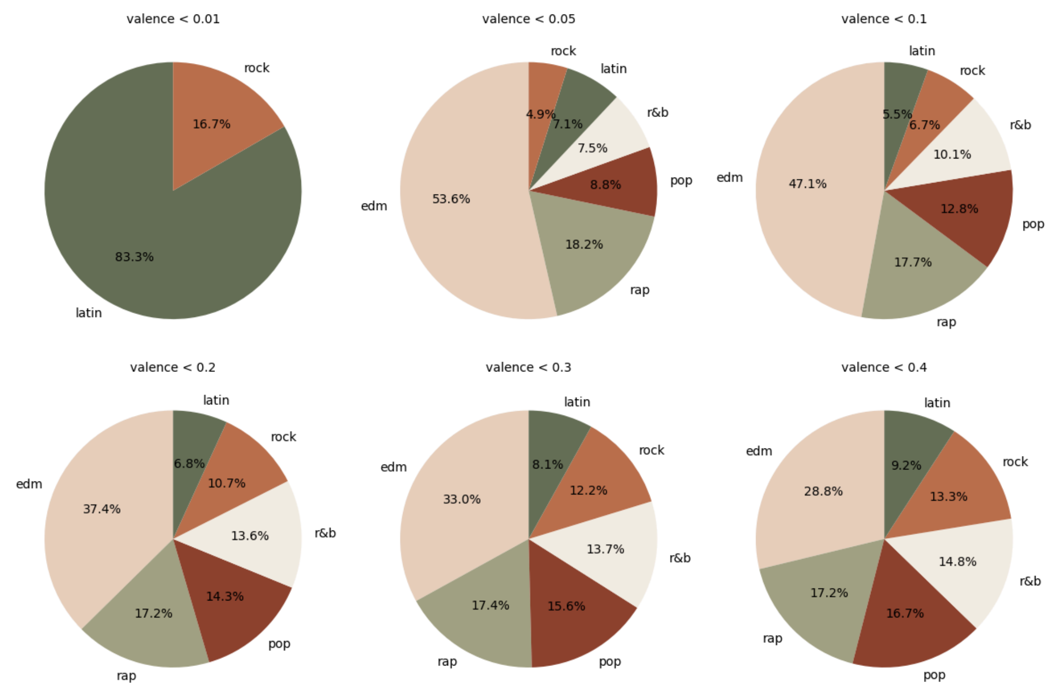
When instrumentalness is above 0.4, edm is most likely.

Valence

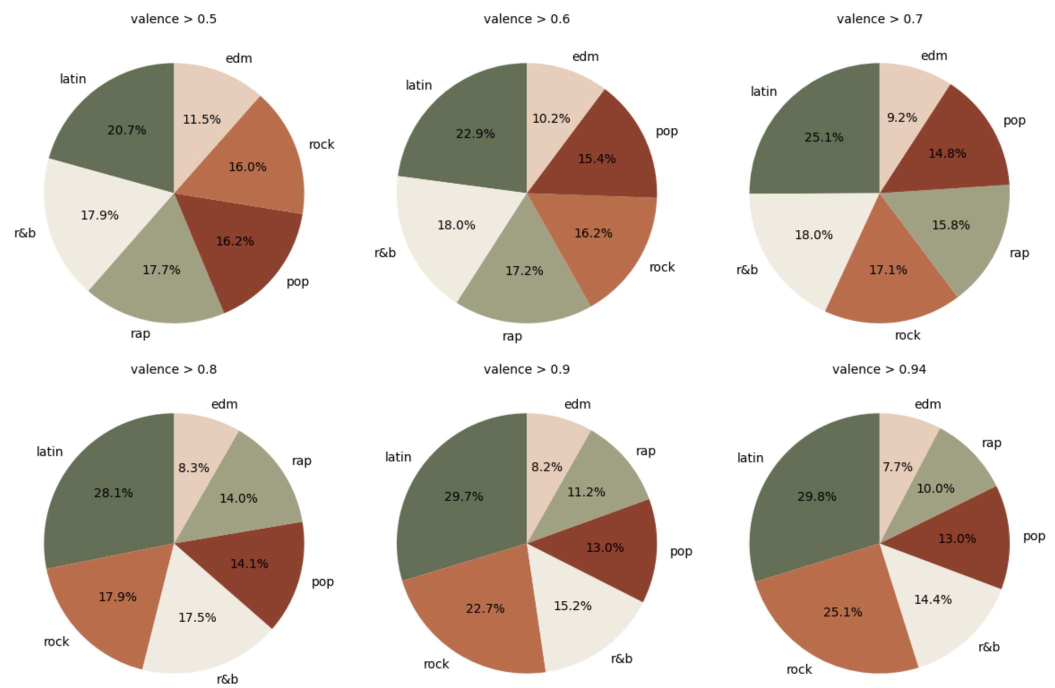
Range Mean: 0.4 - 0.605

Range: 0.00001 - 0.990





When valence is less than 0.1 but greater than 0.01, edm is most likely. When valence is less than 0.01, the only options are latin and rock.

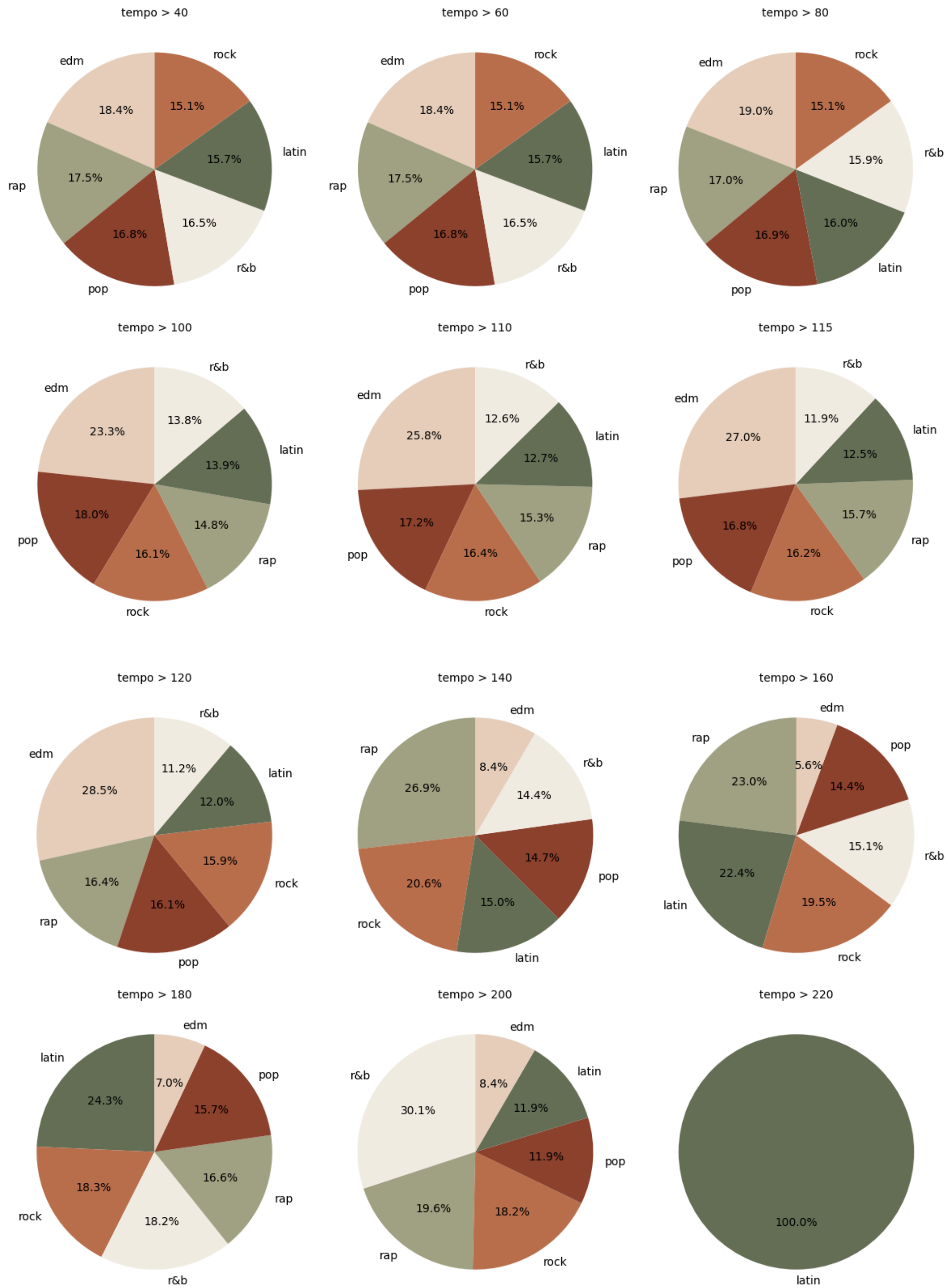


When valence is greater than 0.9, latin and rock are most likely.

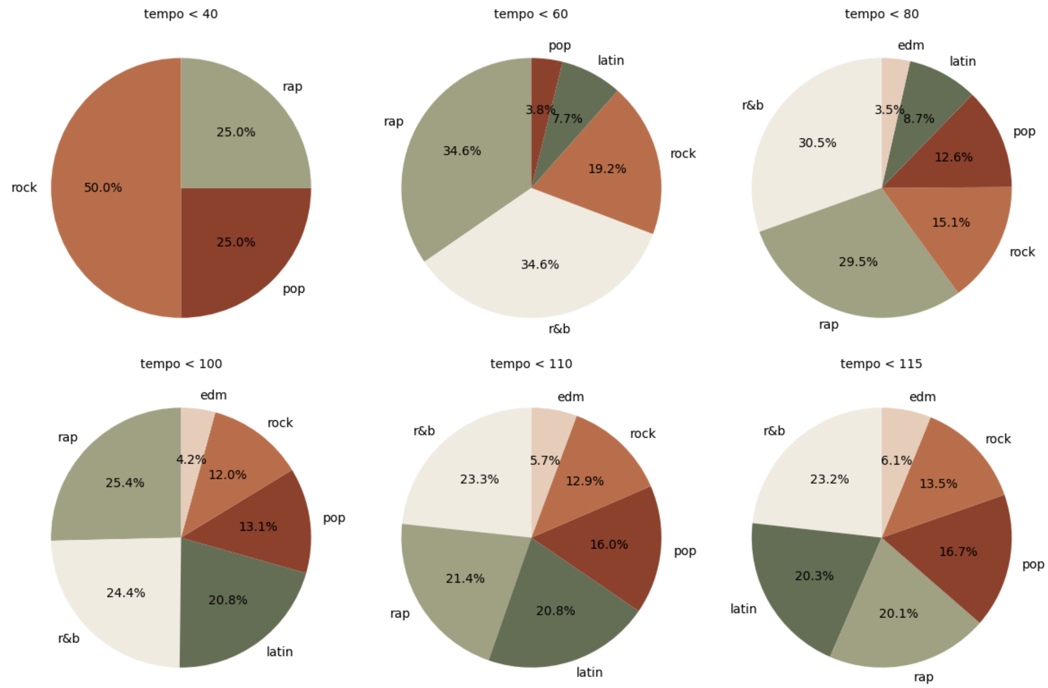
Tempo

Range Mean: 114.222 - 125.768

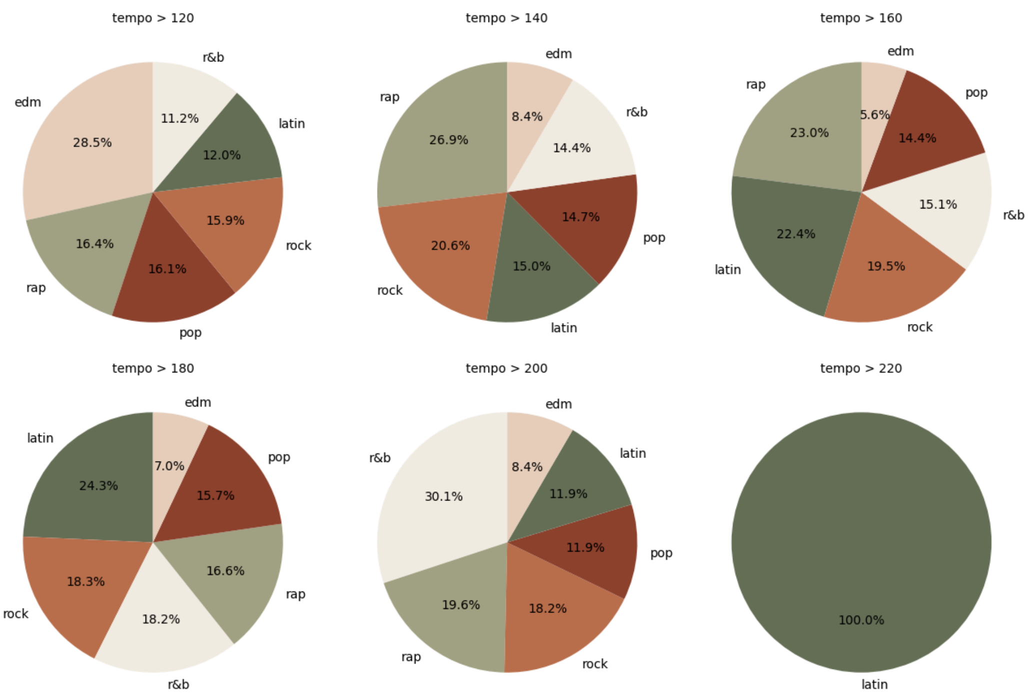
Range: 35.477 - 239.44



When tempo is greater than 220, it is probably a latin playlist.



When tempo is less than 115, edm is not as likely.

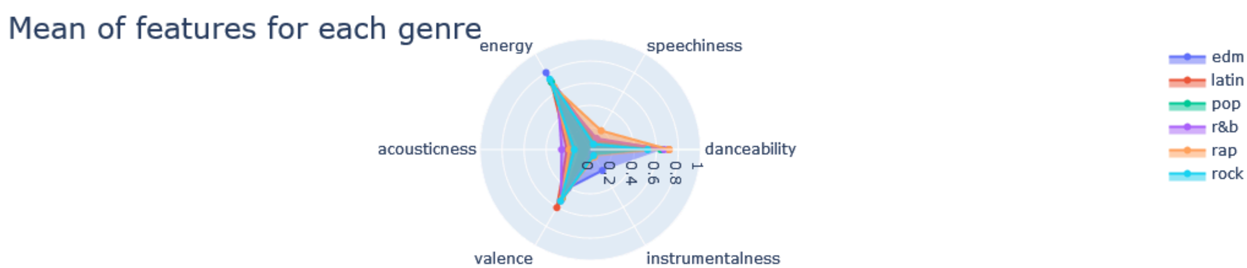


When tempo is greater than 220, latin is the most likely.

Interpretation

By analyzing the distribution of different thresholds of each variable, we are able to determine when specific genres are more or less likely to occur. These pie charts show that key and mode do not have a great effect on genre, but all other variables have at least some effect. By analyzing these charts, we can conclude that at least some of our variables have an effect on the playlist genre, so we should be able to predict the playlist genre with at least some accuracy by fully analyzing the necessary variables.

Radar Chart



Interpretation

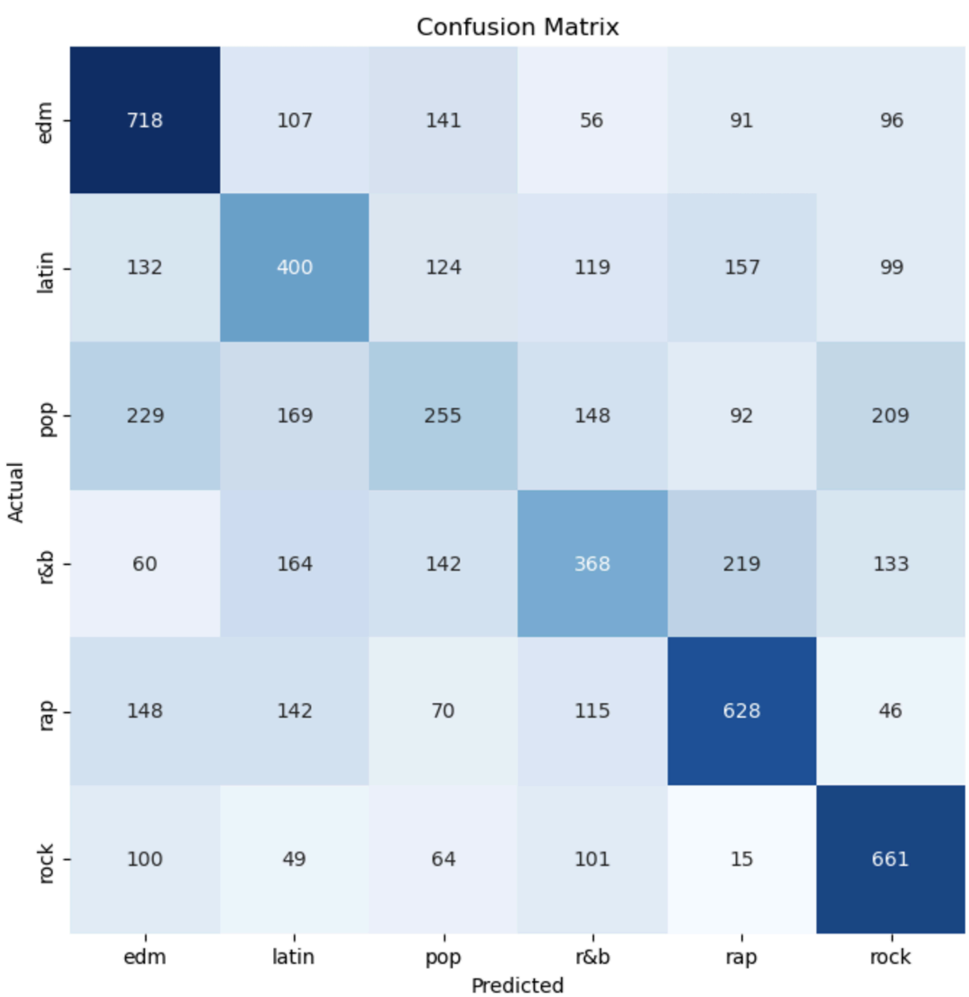
The radar chart compares the mean of each feature across the six genres. To make the chart readable, only features with the same scale of 0.0 to 1.0 were included in this figure. The radius represents the mean value, meaning we can visualize notable differences in means by looking for large gaps between radiuses. For instance, the mean instrumentalness for EDM is significantly higher than other genres. Other features of note are speechiness, where rap has a higher mean, and acousticness, where R&B is visibly higher. This type of figure aids in feature selection by making it easier to compare differences in means between features and then select features that are more distinct between groups. However, the radar chart does lack readability with several categories, especially when the differences between means are not drastic.

Our Results

Five models were developed to predict the playlist genre (playlist_genre): Logistic Regression, Neural Network, Support Vector Machine, K-Nearest Neighbors Classifier, and Random Forest Classifier. All models were trained using these features: danceability, energy, loudness, speechiness, acousticness, instrumentalness, valence, and tempo. Key and mode have been excluding from the training sets.

Logistic Regression

Confusion Matrix:



Classification Report:

Classification Report:

	precision	recall	f1-score	support
edm	0.52	0.59	0.55	1209
latin	0.39	0.39	0.39	1031
pop	0.32	0.23	0.27	1102
r&b	0.41	0.34	0.37	1086
rap	0.52	0.55	0.53	1149
rock	0.53	0.67	0.59	990
accuracy			0.46	6567
macro avg	0.45	0.46	0.45	6567
weighted avg	0.45	0.46	0.45	6567

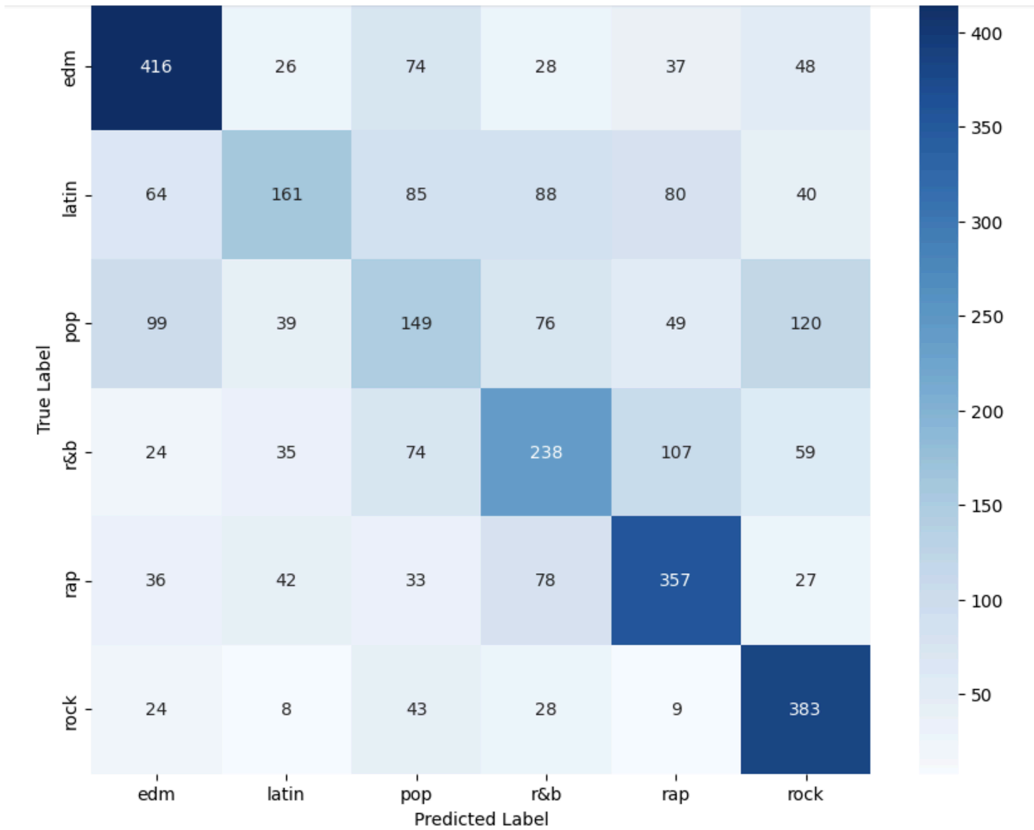
Accuracy: 0.46

Interpretation:

The logistic regression model yielded an accuracy of 46%, reflecting moderate success in predicting playlist genres based on our list of features. The confusion matrix highlights that genres like rock and edm were predicted with relatively higher recall and F1-scores, suggesting these genres have more distinct feature patterns. In contrast, genres like pop exhibited lower predictive performance, indicating overlapping characteristics with other genres. These results suggest that while the model captures some genre-specific trends, additional features or more complex models may be needed to improve classification accuracy.

Neural Network

Confusion Matrix:



Classification Report:

```
Classification Report (test dataset):
              precision    recall  f1-score   support

         edm       0.63      0.66      0.64       629
       latin       0.52      0.31      0.39       518
         pop       0.33      0.28      0.30       532
         r&b       0.44      0.44      0.44       537
         rap       0.56      0.62      0.59       573
        rock       0.57      0.77      0.65       495

    accuracy                           0.52      3284
   macro avg       0.51      0.52      0.50      3284
weighted avg       0.51      0.52      0.51      3284
```

Interpretation:

The neural network model achieved an average F1 score of 0.50 and an overall accuracy of 52%, as shown in the classification report. Among the six music genres, rock and edm exhibited higher F1 scores of 0.65 and 0.64, respectively, suggesting that these genres were more consistently and accurately identified. This can be attributed to their distinct audio characteristics, such as loudness and energy, which the model could distinguish effectively. Similarly, rap also performed relatively well with an F1 score of 0.59, likely due to unique features like tempo and rhythm patterns.

In contrast, pop, latin, and r&b genres had lower F1 scores of 0.30, 0.39, and 0.44, respectively. This indicates that the model struggled to differentiate these genres, potentially due to overlapping feature distributions that make their classification more ambiguous. The confusion matrix reveals additional details about the model's performance. For example, rock and edm had the highest correct classifications compared to other genres, as indicated by their strong diagonal values. However, there is noticeable misclassification between genres like latin and pop, as well as r&b and rap, highlighting the model's difficulty in distinguishing between these groups.

In summary, while the neural network model achieves moderate success in classifying distinct genres like rock and edm, it struggles with genres that share overlapping features. Further refinements in data preprocessing, feature selection, and model design could significantly enhance its performance.

Support Vector Machine (SVM)

Confusion Matrix:

	Pop	**Rock**	**EDM**	**Rap**	**Latin**	**R&B**
Pop	695	116	151	41	111	104
Rock	136	387	114	122	169	105
EDM	210	176	247	178	82	188
Rap	55	154	102	369	229	122
Latin	144	153	61	116	646	48
R&B	108	55	72	118	8	675

The true positives are highlighted in yellow. As seen above, the largest value in each row/column is the true positive value. The model is able to predict the playlist genre with at least some accuracy.

Classification Report:

	precision	recall	f1-score	support
pop	0.52	0.57	0.54	1218
rock	0.37	0.37	0.37	1033
edm	0.33	0.23	0.27	1081
rap	0.39	0.36	0.37	1031
latin	0.52	0.55	0.54	1168
r&b	0.54	0.65	0.59	1036
accuracy			0.46	6567
macro avg	0.45	0.46	0.45	6567
weighted avg	0.45	0.46	0.45	6567

Interpretation:

R&B, latin, and pop all have relatively high precision scores (which means fewer false positives). EDM has a low precision score which means that many instances that were predicted to be EDM were actually a different genre. R&B had the highest recall followed by pop and latin. This means they had fewer false negatives. EDM, on the other hand, had the lowest recall rate so there were many times when there were false negatives. Pop, latin, and R&B had the greatest f1 scores meaning that they had a better balance between precision and recall than the other genres. EDM had the lowest f1 score. Overall, pop, latin, and R&B had the greatest performances. EDM had the weakest performance, most likely because some of its features overlapped with other genres.

This model had a 46% accuracy rate. This shows that the model can predict the genre with some accuracy, though it is not able to predict genre with an accuracy greater than 50%. This is likely because the SVM model might not be able to capture nonlinear relationships within the data and feature overlap may make it difficult for the model to separate genres.

K Nearest Neighbors Classifier

Confusion Matrix:

Row - actual, Column - predicted

	EDM	Latin	Pop	R&B	Rap	Rock
EDM	398	41	99	28	37	26
Latin	65	204	83	66	69	31
Pop	104	76	166	69	36	81
R&B	24	63	74	237	94	45
Rap	49	57	40	79	326	22
Rock	34	10	46	31	8	366

Classification Report:

```
Classification Report (test dataset):
              precision    recall  f1-score   support

         edm       0.59      0.63      0.61       629
       latin       0.45      0.39      0.42       518
         pop       0.33      0.31      0.32       532
         r&b       0.46      0.44      0.45       537
         rap       0.57      0.57      0.57       573
        rock       0.64      0.74      0.69       495

    accuracy                           0.52      3284
   macro avg       0.51      0.51      0.51      3284
weighted avg       0.51      0.52      0.51      3284
```

Interpretation:

The K-Nearest Neighbors model resulted in an average F1 score of 0.51 after tuning the hyperparameters. Certain genres, particularly EDM, rap, and rock, are correctly identified more often with this model, while others such as pop and latin are more difficult to classify. One explanation for this could be that genres such as EDM are more distinct, whereas pop music tends to overlap with many other genres. For the KNN model in particular, it struggles when there are many dimensions/features like in this scenario. In addition, because many features overlap between genres, it is difficult for a model that relies on “clusters” of data points to delineate the different categories. When selecting features for the model, I decided to use backward selection. However, there was no increase in performance when removing features, so the problem remained with having several dimensions. The changes that improved performance were increasing the hyperparameters of the KNN model n_neighbors and weights; specifically, increasing n_neighbors to a higher value (found to be 17) and changing the weighting of nearby points from uniform to being by distance.

Random Forest Classifier

Confusion Matrix:

```
[[1233   97  263   67   91   67]
 [ 145  606  246  204  247   75]
 [ 305  193  517  251  119  266]
 [  64  159  185  722  329  132]
 [ 121  174   85  222 1075   65]
 [  55   42  133  121   30 1144]]
```

Classification Report:

```
Accuracy: 0.537766497461929

Classification Report:
              precision    recall  f1-score   support

         edm       0.64      0.68      0.66      1818
       latin       0.48      0.40      0.43      1523
         pop       0.36      0.31      0.34      1651
         r&b       0.45      0.45      0.45      1591
         rap       0.57      0.62      0.59      1742
        rock       0.65      0.75      0.70      1525

    accuracy                           0.54      9850
   macro avg       0.53      0.54      0.53      9850
weighted avg       0.53      0.54      0.53      9850
```

Interpretation:

These results tell us that the Random Forest Classifier model has an accuracy of 54%. We can see the distinction of accuracy between the genres. Genres like rock and edm are pretty high accuracy compared to the rest. This can logically be assumed as these genres have distinct features like energy or loudness, which can differentiate them from others. Genres like pop, latin, and r&b have lower precision and recall, which can tell us that they probably have overlapping feature distributions. To improve this model, we could incorporate additional features. We could also work on hyperparameter tuning.

Our Interpretation

Our Accuracies:

Random Forest Classifier	54%
K-Nearest Neighbors Classifier	52%
Neural Network	52%
SVM	46%
Logistic Regression	46%

The accuracy results for the Spotify songs dataset tell us a couple of things. We can see that not all the models performed the same. The Random Forest Classifier achieved the highest accuracy which tells us that ensemble models might be the best at seeing and using the non linear relationships with the features in the data (random forest classifiers use multiple decision trees). The K-Nearest neighbors and the neural network both achieved 52% which can indicate that these models had moderate success when identifying patterns. The Logistic regression and SVM models 46% accuracy tells us that the data had possibly overlapping classes or was not linearly separable.

Our Process

Our approach to this project was methodical and collaborative, starting with a detailed analysis of the dataset before moving on to model creation. The dataset contained 28,356 songs across five playlist genres — pop, R&B, Latin, rap, and EDM — and included features like danceability, energy, key, mode, loudness, speechiness, acousticness, instrumentalness, valence, and tempo. To understand how these features related to playlist genres, we first focused on visualizing the data using various figures and charts.

We used scatter plots (created by Senait) to analyze the relationships between features and genres. To make the data more interpretable and account for overlapping points, we jittered the scatterplots and added a mean line to highlight overall trends.. Histograms (created by Sachi) helped us examine how each feature was distributed and whether there were skews in the data. Bar charts (created by Sarayu) allowed us to compare average values, like tempo and valence, across the different genres. Pie charts (created by Eleanor) showed how genres varied when specific features exceeded thresholds, offering a deeper look into variable-genre relationships. Finally, radar charts (created by Snarr) provided a way to visually compare multiple features at once, with each genre represented in a distinct color. These visualizations were essential for identifying which features mattered most for predicting genres.

During this process, we realized that not all features were equally important. For example, the variables key and mode showed minimal variability across genres and had little predictive value, so we decided to exclude it to simplify our models and improve efficiency. In contrast, features like speechiness, acousticness, and instrumentalness showed clear differences across genres, making them crucial for our analysis.

Once we had a clear understanding of the data, each team member focused on their specific responsibilities. Senait implemented the Logistic Regression model to quickly identify the strongest predictors of playlist genre. Sachi worked on the Neural Network model, leveraging its ability to detect complex, non-linear patterns in the data. Eleanor developed the Support Vector Machine (SVM) model, experimenting with different kernels to handle non-linear relationships. Sarayu implemented the Random Forest Classifier, which was particularly effective at capturing non-linear relationships and handling noise. Snarr focused on the K-Nearest Neighbors (KNN) model, fine-tuning the k-value to optimize performance.

Collaboration was key throughout this project. We had regular communication to discuss our progress, share challenges, and ensure that everyone was on the same page. Responsibilities were divided based on each person’s strengths and interests, which allowed us to work efficiently and produce high-quality results. We used shared tools like Google Drive and Jupyter notebooks to integrate our work, and we frequently reviewed each other’s contributions to maintain consistency and accuracy. The final report was a team effort, with everyone contributing to the sections that aligned with their tasks.

In the end, this process showed how valuable a well-organized and collaborative approach can be. Using figures to explore the data before building models allowed us to focus on the features that mattered most, like speechiness and acousticness, while dropping less relevant ones like key and mode. Each team member's contributions, combined with our regular collaboration, ensured the project was both thorough and efficient. Moving forward, there's potential to enhance our work by incorporating external datasets or fine-tuning our models further, which could lead to even better predictions.

What We Learned About Our Data

In our exploratory data analysis, the figures we created helped us identify which features differed across genres. We learned that energy, speechiness, acousticness, and instrumentalness have significantly different mean values in different genres. Examples of this included EDM's low acousticness and high instrumentalness, and rap's high speechiness. These were features we decided to use in many of our models. In contrast, the figures also helped identify features like key and mode as being similar across genres and thus unlikely to be helpful to include in our models. The histogram and bar chart were particularly useful in visualizing trends within our dataset.

The weaknesses of certain models were highlighted in this process. The logistic regression and support vector machine models had their performance limited by the fact that the genres' features have heavy overlap with each other, making them difficult to separate with lines or vectors. The accuracy of the K nearest neighbors model was limited by the fact that as more features were added to try to improve performance, dimensionality increased, which hindered it because it is distance-based.

The neural network, K nearest neighbors, and random forest classifier models displayed the same pattern when identifying genre. All three were better at identifying EDM, rap, and rock when compared to the other 3 genres (pop, latin, and R&B). This could suggest that these music genres have more distinct qualities. This was also noted in the exploratory figures.