# Spotify Music Data Exploration


## Andrés De La Rosa, Anand Patel, Lina Yang



04/16/2021| W200 |Sec 12: Fri 4:00-5:30pm PT


## INTRODUCTION

Music plays an important role in the lives of people all around the world, evident with Spotify's 345 million monthly active users and 155 million subscribers. People's music preferences vary, and they listen to music for a variety of reasons. While music journalism publications, such as Pitchfork, are considered tastemakers in determining what songs are popular or are good, everyone listens to music differently. More often than not, a hit is not always correlated with a critic's score. Among the three of us, our music choices are diverse. They can be measured in various ways, including by artist, album, duration, and audio characteristics such as danceability, energy, and acousticness (to name a few). Among the three of us, do our personal tastes in music converge on similar audio and song characteristics? And, as we observe the characteristics of 2017 song tracks and their Pitchfork ratings, is there any correlation between audio and song features with good and bad ratings? Additionally, of those correlated features with Pitchfork's highest-rated songs, can we correlate them with our personal preferences, revealing whether critics are connoisseurs of songs or are just an opinion?

## HYPOTHESIS

We predict that among the three of us that we have very different tastes in music, but the music features from the top rated music will converge with our top music features.

Based on each of our Spotify liked songs library data set, we hypothesize that Spotify audio features we differ most on would stratify our Spotify liked songs libraries uniquely and allow us to retrieve album recommendations from Pitchfork based on the summary of audio features for tracks encompassing our stratified audio feature the most. For example, a high danceability song for each team member might look very different based on all the other audio features of our tracks. For some team members, a highly danceable track might be very happy (valence) and for others it might be something more instrumental, like some indie rock music. As a result, we believe that looking at subsets of our libraries based on highs for danceability and instrumentalness, and characterizing the subset based on its averaged audio features with weights on importance, would allow us to find Pitchfork reviewed albums that we might individually enjoy based on audio features prevalent in our libraries. In short, if we're looking for danceable or instrumental new music, then our remaining audio features among our most danceable and instrumental liked tracks should help us find similar album recommendations from Pitchfork.

Additionally, it can be insightful to see the scores, genres, and reviews of the Pitchfork albums that are returned for each team member to understand how aligned our tastes our with music journalists, what genres we might enjoy, and even some common themes that come up in the reviews for Pitchfork albums we might like.

## DATA SOURCES:

1. Spotify: 400 songs from each of our personal libraries
2. Pitchfork (P4K) reviews via Kaggle.com (https://www.kaggle.com/ermoore/pitchfork-reviews-through-12617)
3. Spotify music features of the Pitchfork reviewed songs

## DATA

Below is a table of the Spotify data:

| Data Feature | Definition | Type |
|---|---|---|
| username | The owner of the Spotify library. It will either be Anand, Andrés, or Lina. | string |
| uri | Spotify's unique identifier for the track. Entering this id into Spotify's search will bring you directly to that specific track. | object |
| name | song name | string |
| popularity | The popularity of the track. The value will be between 0 and 100, with 100 being the most popular.The popularity of a track is a value between 0 and 100, with 100 being the most popular. | integer |
| added_at_playlist | The date the song was added to user's playlist | date |
| duration_ms | Duration of song in milliseconds. | Integer |
| is_explicit | If a track is considered to be explicit the element will be set to true, otherwise false. | Boolean |
| album_name | Name of album the song belongs to. | string |
| album_release_date | The date of when the album was released. | date |
| artists | The artist that sang the song | string |
| artists_id | Spotify IDs for the artists | string |
| 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. | float |
| energy | 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. | float |
| 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. | integer |
| 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 typical range between -60 and 0 db. | float |
| 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. | Integer |
| speechiness | Speechiness detects the presence of spoken words in a track. The more exclusively | float |

| | | |
|---|---|---|
| | 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. | |
| acousticness | A confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic. | float |
| instrumentalness | 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. | float |
| 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. | float |
| 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). | float |
| 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. | float |

Below is a table of the Pitchfork Reviews data:

| Data Feature | Definition | Type |
|---|---|---|
| album | Name of the album. | String |
| artist | Artist of the album. Usually multiple artists are listed as "Various Artists" | String |
| best | 1 or 0 if the album received a "best new music" label for the current year of the release. Applies only to first released albums. | Boolean |
| date | Date of the album review. | String |
| genre | Genre of the album. | String |
| review | The music journalist's entire detailed review of the album. | String |
| score | The score the album received from 0 through 10. Higher score means more acclaimed. | float |

## DATA CLEANING & SANITY CHECKS

### Spotify Data

The data from Spotify was obtained by calling the Spotify API to retrieve the data from each of our top 400 saved songs. The initial datasets included a list of authors, songs, and their features. The Spotify data was relatively clean in regards to having values in each of the features for each of the songs. We had to add a new column feature to each of our data sets to identify whose data is whose. Then we joined the authors and songs datasets together using the row indexes as the unique identifier. Finally, we unioned each of our 400 songs to one dataset. The resulting dataset has 21 columns and 1,200 songs. Some artists and songs are in non-English characters, but are showing up correctly in the data frame in the original language characters. This also should not cause any issues with our statistical analysis because we are not performing any natural language processing.

### Pitchfork (P4K) Data

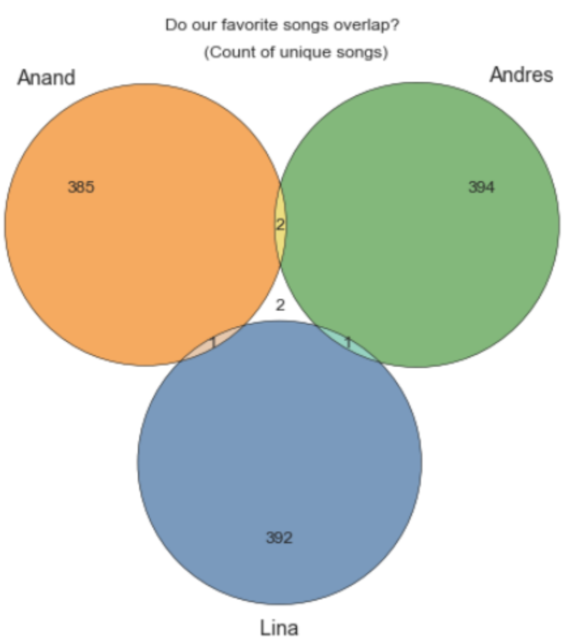
The P4K data was obtained from kaggle and it contained reviews going back to the start of the website.

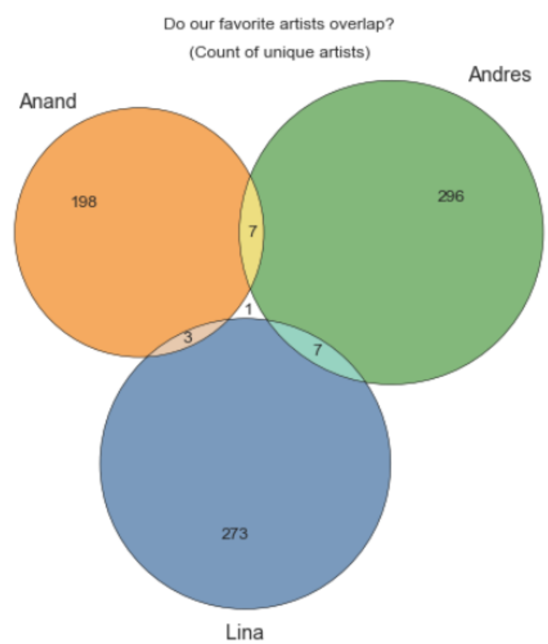
We downsized the dataset to only have the 2017 album reviews spanning the time frame between Dec 6th 2016 to Dec 6th 2017. There were 1256 total albums or rows in this dataset, and every field was populated properly. For each album in this dataset, we used the Spotify API to search for the album name and artist name listed and bring back the album_uri if available on Spotify. Since some albums were not available on Spotify, our P4K dataset's final size included 1009 albums or rows. For each album, we used the Spotify API to retrieve each album's tracks, along with the audio features minus 'mode' and 'key'. This brought back 12,298 tracks. To determine the Spotify audio features of an entire album, we averaged the individual audio features of its tracks and added these variables to its album row in our 1009 P4K 2017 albums data. For album_duration we summed the duration of each of its tracks, and for is_explict we summed the number of explicit tracks in the album. Our final P4K 2017 dataset contains all the variables from the P4K dataset on kaggle for 1009 albums reviewed in our study time span, and also the Spotify audio features, album_uri, album_name, and artist_name found via our methods and API search.

## RESULTS

There are many ways we can test our hypotheses. It can be by music features, genre, artists, and songs (just to name a few). Spotify's API does not allow us to download music genres (it has been a known broken issue in the API for a few years), but we can explore our music preferences quantitatively by the different features as listed under our Data section.

***Do our favorite music and artists intersect within each of our top 400 songs?***



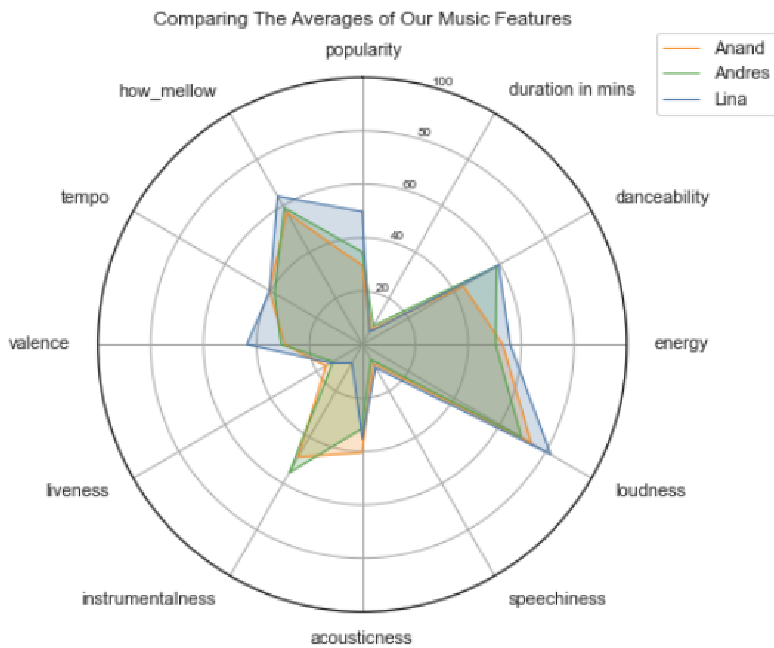




Based on each of our 400 song libraries, it looks like we have very little overlap in the music we listen to. For example, of the unique artists we have in our libraries, Anand and Andres only have 7 artists in common, while Lina and Anand only have 3. Among the three of us, Andres has the most artists that intersect with Anand and Lina. It is also worthy to note that Andres has the largest volume of unique artists in his library followed by Lina and then Anand, which most likely contributed to the probability of why Andres had more intersecting music with Anand and Lina.

The one artist that the three of us have in common is Bicep. The artists that Anand and Andres had in common were Against All Logic, Chromatics, David Bowie, Floating Points, Jon Hopkins, and Nicolas Jaar. The artists that Lina and Anand had in common were Aphex Twin, FKA twigs, and Radiohead. The artists that Andres and Lina have in common are Arcade Fire, Four Tet, Joywave, Tame Impala, The xx, [Howling, RY X, Frank Wiedemann], and [ODESZA, Zyra]. By initial research, the artists that we seem to intersect most on are rock or electronic artists. When we look at the count of unique songs the volume of the output increased, but resulted even less in commonalities. While the three of us intersect on two songs, when comparing one to another, our common songs are 1-2 per pairing.

It seems like our hypothesis on a high-level based on unique artists and songs is proving to be true.

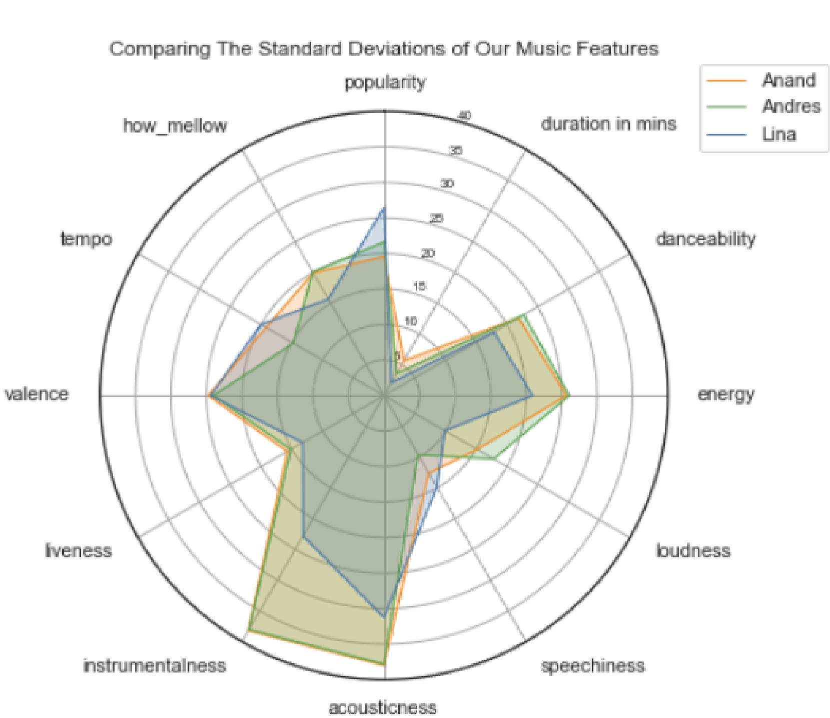
***What is the relative weight of our music preferences across quantitative variables?***




While our music by artist and song title may not have much commonalities, it would be prudent to evaluate our music preferences on a quantitative scale based on the music features Spotify offers. As notated in the Data section some of the music features have varying scale. This can prove to be hard to compare without transforming the data somehow. In the below use case, we have decided to do a simple linear transformation of each song feature value to convert to a 0-100 scale.

Then we either did a simple average (mean) to show a "general" idea of what type of music we like to listen to and then a standard deviation of the song features to show the variety of music we like to listen to.

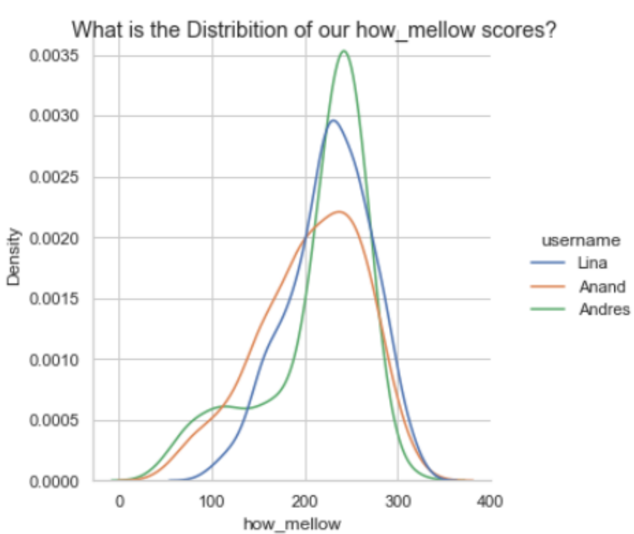
When taking a look at the simple average of our music features our general music preferences show some weight towards certain music features. For example, we all have a higher average score for




danceability, loudness, and mellowness (which means we prefer less mellow music; we will go into more depth later). Anand and Andres score higher in average instrumentalness than Lina, and Anand scores higher in average acousticness with his music.

Parallelly, we can look at the features based on standard deviation of each music feature. As we can see in the chart, this looks slightly different than the simple average of our music features. Anand and Andres both have a high standard deviation when it comes to instrumentalness and acousticness which can indicate that songs based on those audio features vary quite a bit, while with Lina, she stays within a standard deviation pocket across most of the music features with slightly higher standard deviations with acousticness and popularity compared to other features.

***How mellow do we like our music?***



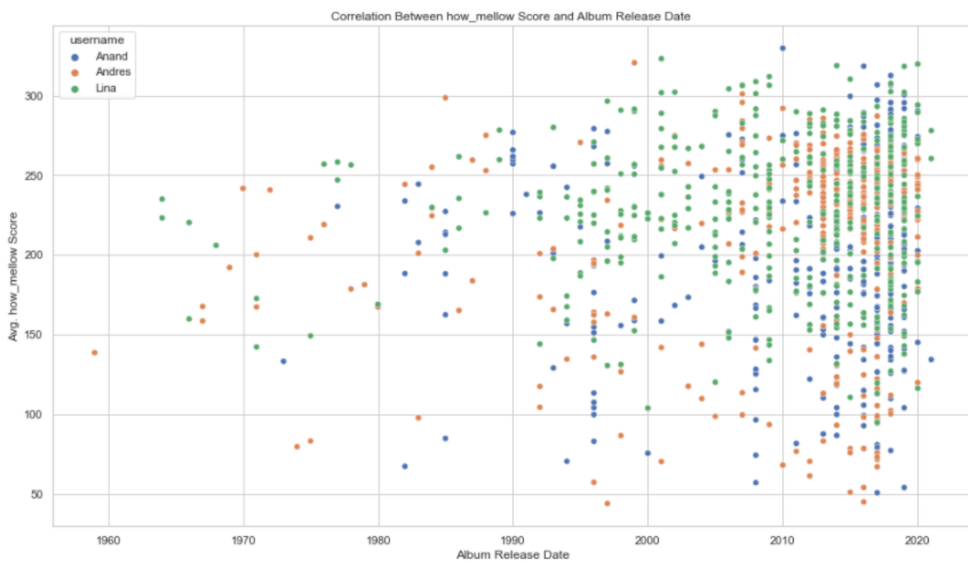

If we have discovered anything so far in this report is that music is complex and can be measured in many different ways. To reduce the features to one variable, we tried to come up with a calculation of a set of variables that can help determine the type of music we like. Inspired by a source we found online, we attempted to use this new calculated variable to see if this gleans more into what type of music we like. The author of this article, called it a boringness score. We decided to call the variable *how_mellow* to be less arbitrary. The equation to calculate how_mellow is as follows:

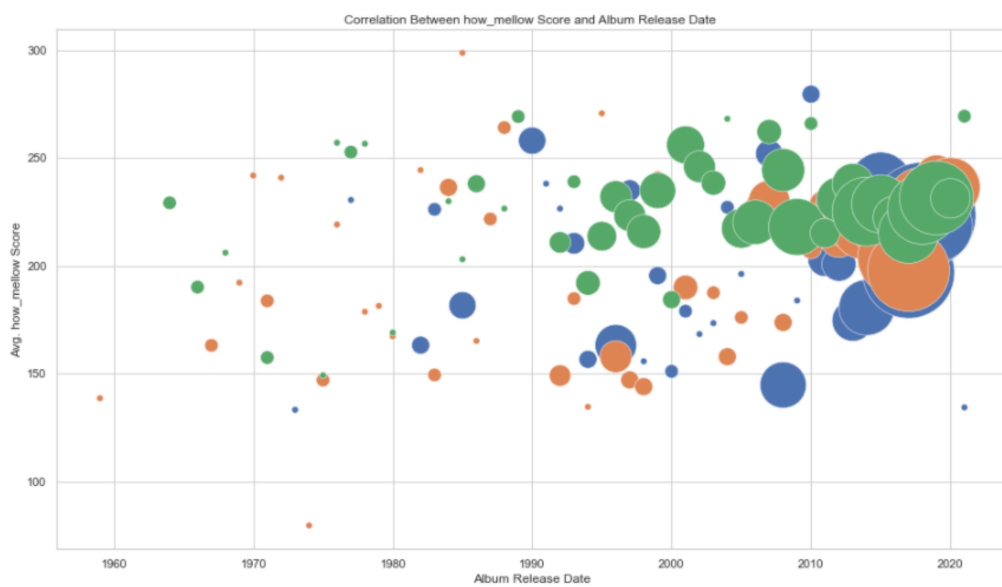
$$how\ mellow\ =\ loudness\ +\ tempo\ +\ (energy\ *\ 100)\ +\ danceability\ *\ 100$$

The lower the score the more mellow the song is and the higher the score the more upbeat and lively. When we look at the distribution of our music against this score. The distribution of all of our how_mellow scores skew right. The majority of our scores concentrate on the higher end on the scoring. Of the three, Anand's scores are the lowest and have a wider spread. Indicating that Andres and Lina are listening to more lively music in their libraries, while Anand prefers more mellow tunes.

### *How does the mellowness of our music correlate with time intervals?*

It's a known fact among music historians that music style changes over time, and with Spotify's immense library we can enjoy music from many generations. With our general music mellowness skewing right towards more lively music, we decided to see if any interesting trends come out compared to album release date. When comparing the frequency of songs in our libraries we see that a large frequency of our music was released within the past decade and much of that music stays within the mellowness score of 150 to 250. We see that Anand's music was mainly released after 1980, while Andres and Lina have larger range of songs from different decades.

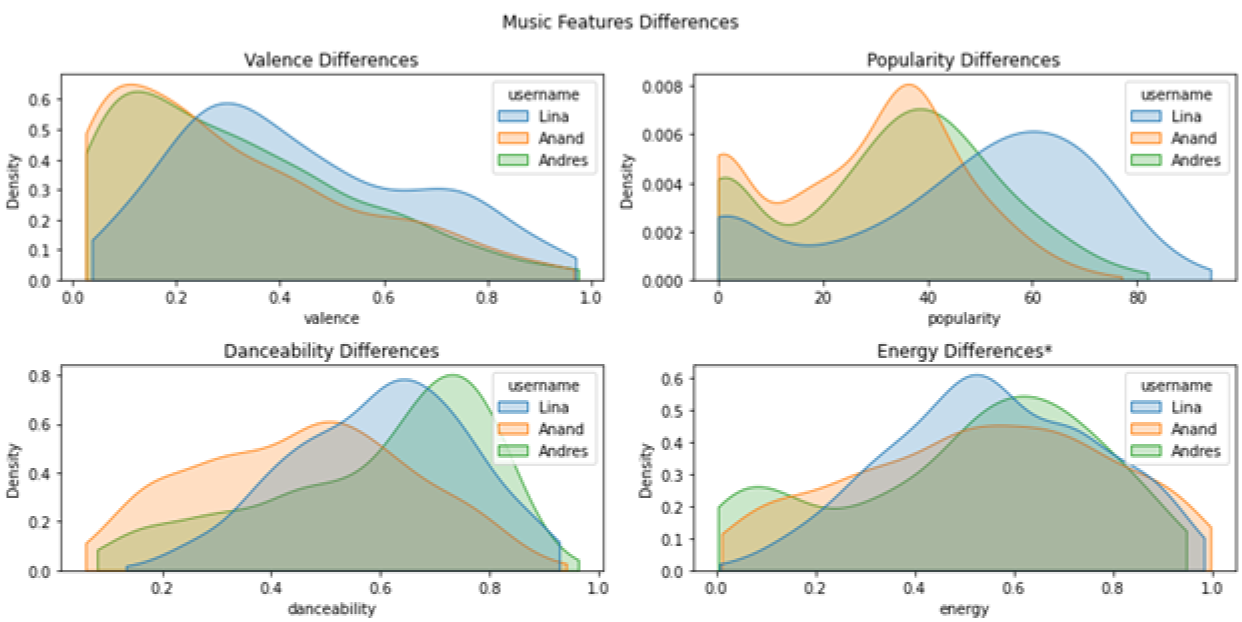


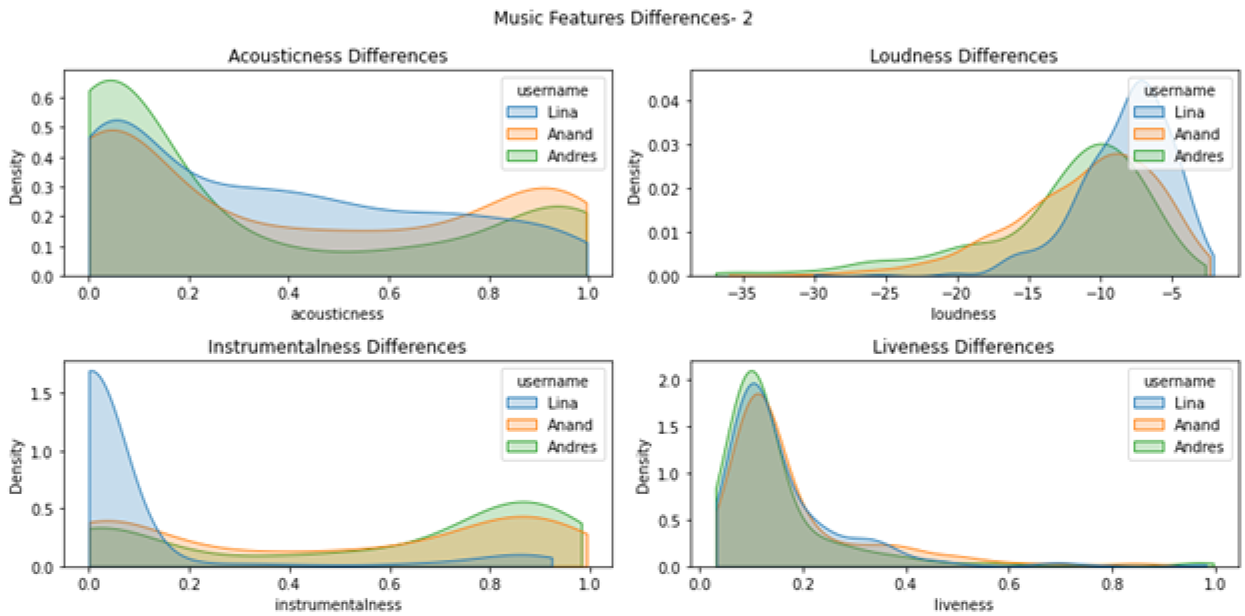
***How different are our team members' musical preferences?***

The type of music that we listen to may be unique for each of us. However, there might be some general similarities between them. Our initial hypothesis is that each one of us would have, on average, different preferences of style from our 400 songs since we have a number of distinct artists and 1,200 other tracks that have different characteristics. Using music features provided by the Spotify API, we proceeded to compare how different were our playlists. Most of these features were continuous variables, so we decided to analyze their distributions at first glance.

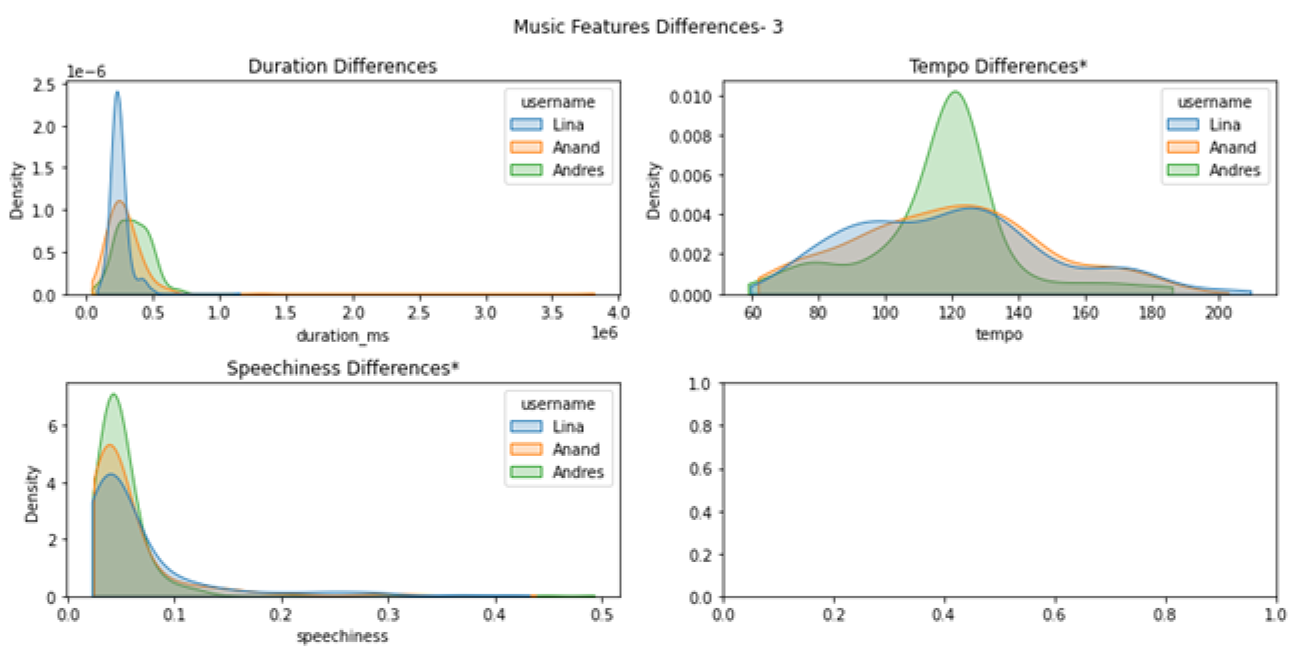
Since these musical features' distribution does not appear to be normally distributed, as we can see from the charts below,  we analyzed them with a non-parametric distribution. We used a non-parametric test to test if they were statistically different.




Spotify defines our songs' happiness as valence, which is a measure from 0 to 1 describing the musical positiveness conveyed by a track. From the chart above, we can see that Lina has a more significant proportion of her songs with more considerable valence and danceability. Also, she listens to more popular songs than Anand and Andres. Regarding energy, which is a measure from 0 to 1 that represents a perceptual measure of intensity and activity, it seems that this variable is not very similar between users, but, according to the Kruskal Wallis test, this musical feature is not statistically different among Lina, Andres, and Anand.




Lina seems to be more consistent around her preferences for acousticness. Anand and Andres prefer, overall, less acoustic music since their tracks' distribution is concentrated between 0 and 0.4, indicating that they like less acoustic music. In terms of loudness, instrulmentalness and liveness (the presence of an audience in the song), distributions appear to be different, which is then confirmed by the Kruskal-Wallis test, where for all these variables, the null hypothesis is rejected.




Regarding duration, Andres prefers longer songs than Anand and Lina, who inclines toward shorter songs. Finally, tempo and speechness are two of the three musical features that each user has in common. Spotify defines speechness as the presence of spoken words in a track, and if the value is more than 0.66, then it will likely be rap or a podcast, here we can see that most of the tracks in the sample are below 0.33. Andres seems to like the tempo of its tracks between 100 and 140, as we can see from the distribution chart. However, according to the statistical test, these differences are not enough to say that it is unlikely.

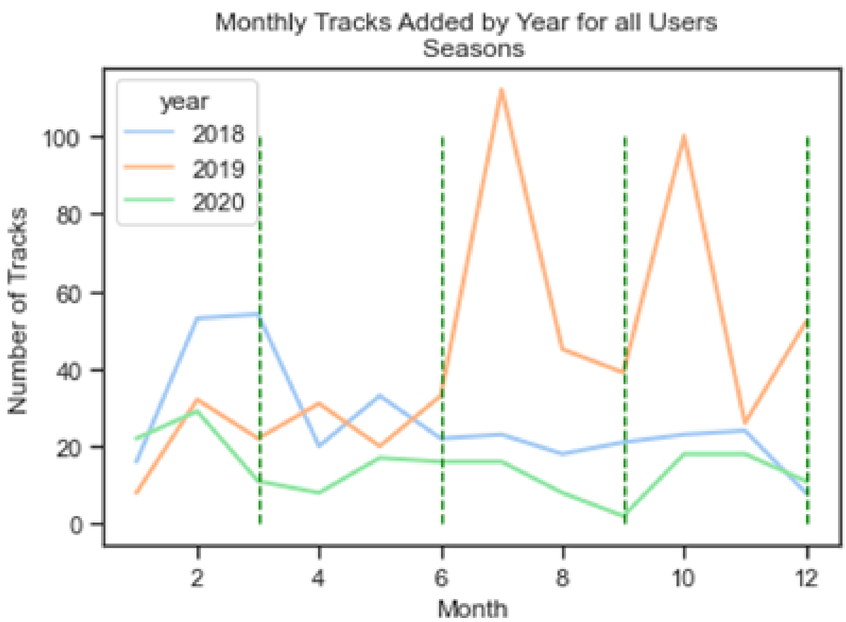
To assess the differences between these variables, we proceeded to perform a Kruskal Wallis test, for each one of the music features, under the following null hypothesis:

*“The Kruskal-Wallis H test (sometimes also called the "one-way ANOVA on ranks") is a rank-based nonparametric test that can be used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable.”*

| Music Feature | Username with min | Username with max | Kruskal Wallis P-value | Is it different? |
|---|---|---|---|---|
| Danceability | Anand | Andres | 0.000 | True |
| Energy | Andres | Anand | 0.109 | False |
| Loudness | Andres | Lina | 0.000 | True |
| Speechiness | Lina | Andres | 0.106 | False |
| Acousticness | Anand | Lina | 0.001 | True |
| Instrumentalness | Lina | Anand | 0.000 | True |
| Liveness | Andres | Andres | 0.000 | True |
| Valence | Andres | Andres | 0.000 | True |
| Tempo | Andres | Lina | 0.264 | False |
| Popularity | Lina | Lina | 0.000 | True |
| Duration_ms | Anand | Anand | 0.000 | True |

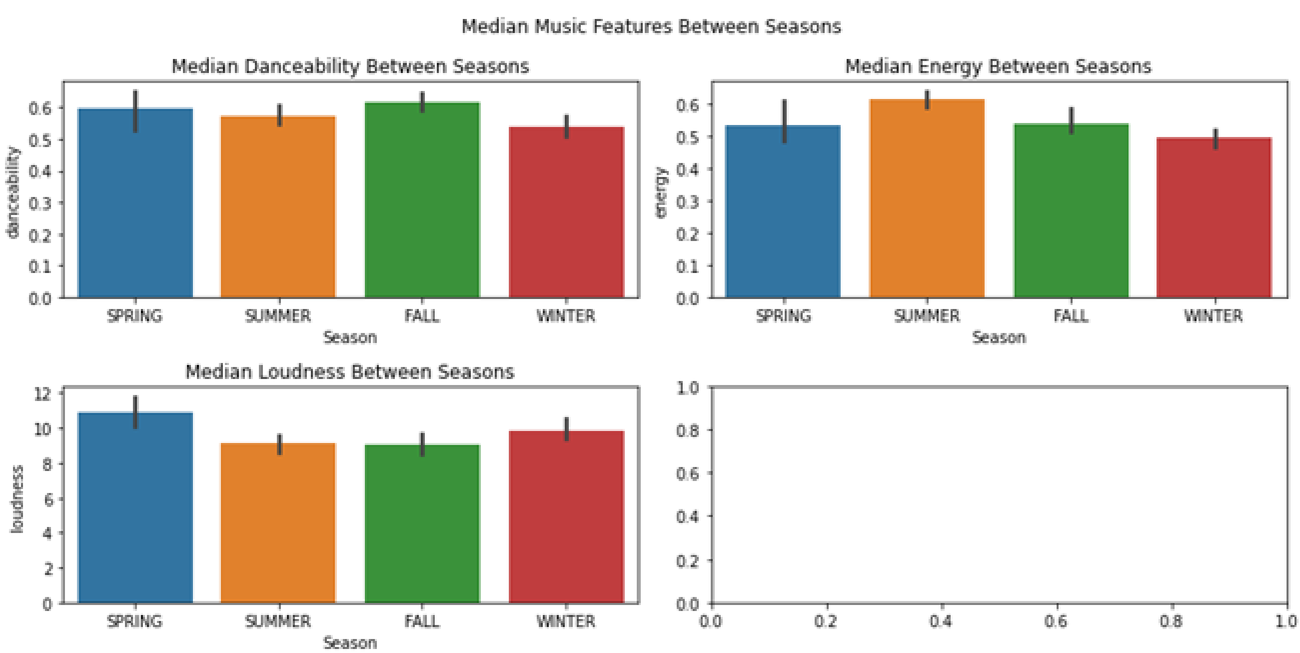
***In which period of the year, we tend to save music into our Spotify library? Do the characteristics of the songs saved correspond to the season?***

Anand, Andres, and Lina are active Spotify users. The three of them have been consistently adding songs for over three years, as shown in the chart below. Thus, we would like to know the features of the music that they tend to add over the seasons of the year. Our initial hypothesis is that they add music with more danceability, energy, and loudness in the summer.




This might change to music with lower danceability, energy, and loudness in the winter because of the lack of sunlight due to the days' shortening. Of course, this will depend individually, but let's see what the data tells us about that.

To answer this question, we looked over each of these variables grouped by season and found that the danceability, energy, and loudness of the tracks added are different between each season after we performed the Kruskal Wallis test.




However, as we can see from the graphs, danceability, energy, and loudness are different in other seasons. The median danceable track added on fall is more significant than on other seasons. The median energy of a track added on summer is greater than another season, meaning that our hypothesis could be correct for this variable. Finally, the median loudness of a track added on spring is more remarkable than other seasons.

We can conclude from our analysis that in summer, we tend to add songs with more energy. Other interesting variables, such as valence, liveness and tempo weren't analyzed because there wasn't enough variability to conclude that the tracks added would differ from season to season.

| Music Feature | P-Value | Is it different? |
|---|---|---|
| Danceability | 0.0010 | True |
| Energy | 0.0000 | True |
| Loudness | 0.0000 | True |
| Liveness | 0.5903 | False |
| Valence | 0.0947 | False |

## Pitchfork Data Analysis

### Pre-processing the P4K Data.

In our P4K album dataset, we apply a "group_label" to capture how each album compares to the average score and average popularity (7.29, 23.7%) of all the albums in the 2017 P4K dataset. This helps label each album as favorites (high score, high popularity), underrated (high score, low popularity), overrated (low score, high popularity), and ignorables (low score, low popularity).

For this analysis we normalize popularity, loudness, and tempo to values between 0 and 1 in the user library data and the Spotify album data by dividing by 100, -60 dB, and 210 bpm respectively. This ensures that we can compute a vector of a user library's audio features with equally scaled dimensions from 0 to 1. This user library vector will be used to compute the distance between library audio features and a P4K album, which returns a scalar value that can represent the album's recommendation characteristics. In the user library vector, we compute the average value of each audio feature in the library dataframe $V_n$ and also a weight for this audio feature $W_n$. The weight is determined from the value's standard deviation away from its mean value. For example, if Anand's library tracks had some valence average value with a small standard deviation then it would mean that most of the tracks are clustered around this average value so its weight should be higher since it seems like Anand really likes valence around this value. Conversely, if the standard deviation was higher then the weight could be lower since this audio feature does not need to have exactly this value.

$$w_n = 100 * \frac{abs(\frac{n.mean()}{n.std()})}{\sum_{i=1}^{k} abs(\frac{i.mean()}{i.std()})}$$

An album's distance is computed by the formula below. It is essentially a distance vector, but each difference between an album's audio feature $v_{n,\,alb}$ and the user library vector's audio feature value $v_{n,\,0}$ is ultimately scaled by the weight of that audio feature from the user library vector $w_n$. We consider the following audio features:  acousticness, danceability, energy, instrumentalness, liveness, loudness, speechiness, tempo, valence, and popularity. We normalized some audio features earlier to ensure that, for example, tempo differences being an order of magnitude higher than audio features from 0 to 1, simply based on how tempo is measured, will not blow up the distance.

$$album's\ distance = \sqrt{\sum_{n=1}^{k} w_n (v_{n,\,alb} - v_{n,\,0})^2}$$

<Why high danceability and high instrumentalness? Because our music feature analysis above shows these features have different densities for all 3 users for danceability and between Lina and Andres & Anand's libraries for instrumentalness.>

From the earlier analysis on audio features, we saw that danceability and instrumentalness  have different densities for all 3 users for danceability and between Lina and Andres & Anand's libraries for instrumentalness. Therefore, we will find album recommendations for high danceability and high instrumentalness. Hence, we will generate user library vectors for each user's top 80% quartile albums by danceability and intrumentalness. The P4K albums will have their distances computed, sorted from lowest to highest distance, and returned as high danceability and high instrumentalness recommendations for P4K 2017 albums for each user.

We generate 3 datasets on all the P4K albums ranked from most to least recommended based on each album's distance from the user's library recommendation vector. Each dataset has a column on distance and a column for the username that this recommendation belongs to.

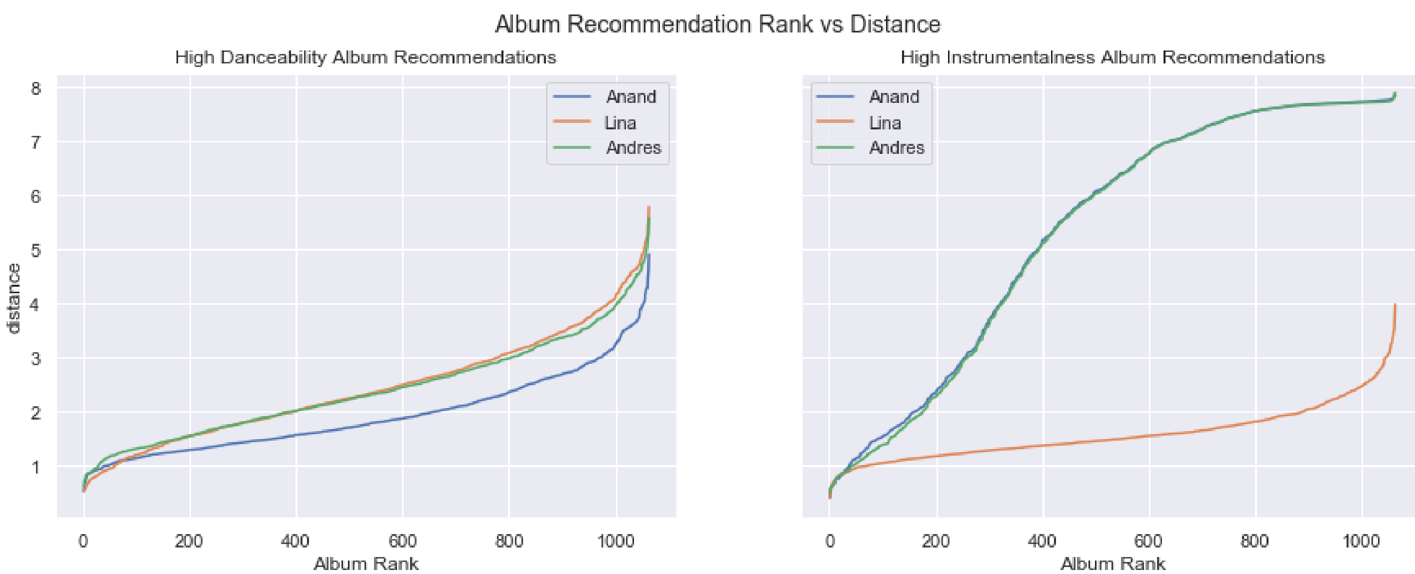
We grab the top 100 recommended albums from each user, from their high danceability and high instrumentalness datasets, and create 2 combined datasets, 300 rows long, of each user's top 100 recommended P4K 2017 albums based on high danceability and high instrumentalness.

Extra variables to P4K: username, distance, group_label. First two are specific to the user this album was recommended to.

### How well do we predict album recommendations?

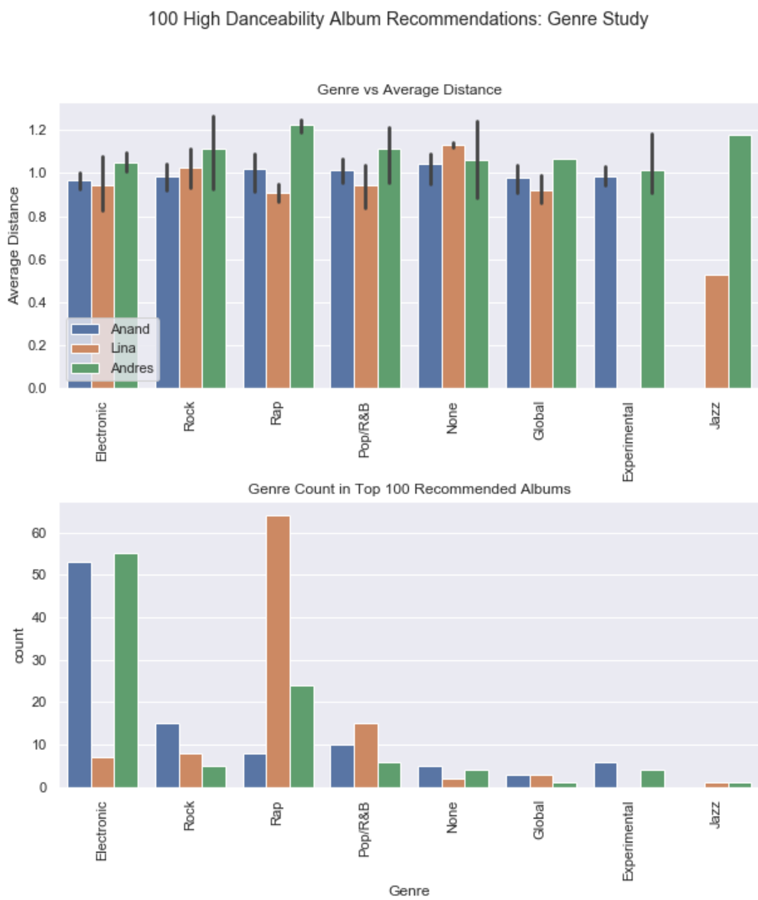
For each of our users, we have the 2017 pitchfork albums ranked by distance to their library search vector from smallest (closest recommendation, highest rank) to largest (furthest recommendation, lowest rank).

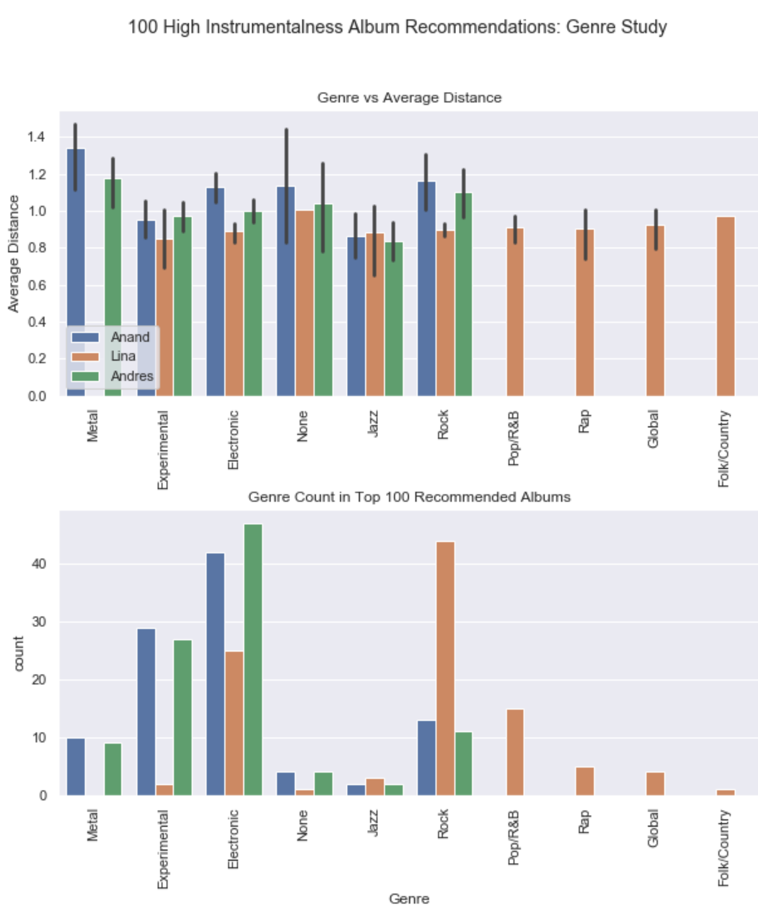
Below we plot our user's distances vs album rank for their entire recommended album lists for high danceability, on the left, and high instrumentalness on the right. We want to know if the distance, or how close an album is to our user's search vector, is a function of the album rank. From the plot we can see that albums that are higher ranked in the recommendations list have much lower distances than lower ranked albums. We would expect this behavior from the sorted list. However, there are key differences in how the distances grow for users in the danceability recommendations versus the instrumentalness recommendations. From the danceability recommendations, we can see that all 3 users have similar values in distances with rank and growth in distance with rank. This indicates that these users were provided with roughly equally confident album recommendations. Since these rank vs distance curves are close, it will be interesting to see if the albums recommended have overlap. In the instrumentalness recommendations, we can see that Lina has much lower distances compared to Anand and Andres, which indicates that her recommendations are more closely aligned with her music taste by audio feature. Anand and Andres' curves overlap, which might indicate overlap in albums recommended. In order to compare recommendations fairly, we restrict analysis to each user's top 100 recommendations to select albums that have reasonably comparable distances before the distance curves diverge for lower ranked albums.




### What does Pitchfork think of our music taste?

Pitchfork classifies its album reviews by the genre of the album, and the album quality by the score of the album. We have also additionally added group_label's to classify Pitchfork albums based on how they rate by score (critical acclaim) and how Spotify keeps track of popularity (audience acclaim). The four categories for an album's group label are: favorites (high score, high popularity), underrated (high score, low popularity), overrated (low score, high popularity), and ignorables (low



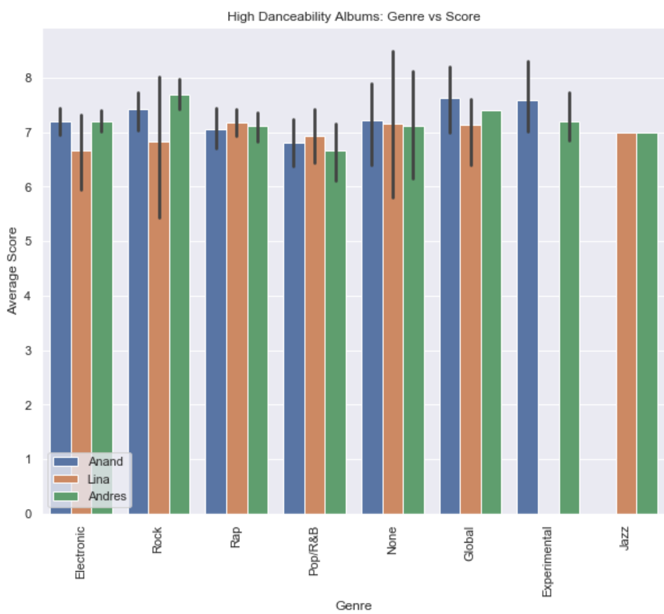

score, low popularity). In this section, we analyze our recommendations based on these 3 classifications from Pitchfork to assess how our music taste, which governs the albums recommended to us, translates to Pitchfork's way of looking at things.

## Genre Analysis

***How many different genres do we listen to and does that have to do with recommendations?***

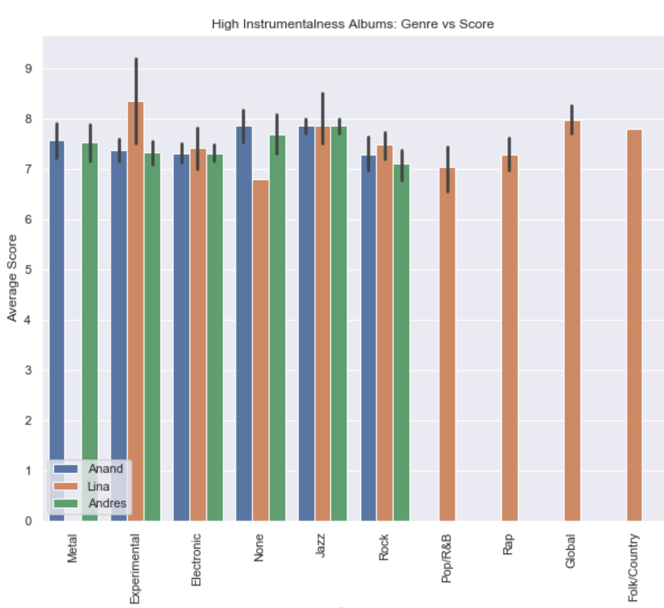
For our danceability and instrumentalness album recommendations, we would like to know how many different genres of music each user was recommended. Among these genres, we also want to know if some genres were more recommended to a user than others. On the left, we can see these results for the danceability recommendations. In these danceable album recommendations, it's clear that Lina's music taste leans most towards rap music compared to other genres and also over the other users since she has over 60/100 rap recommendations. The average distance of Lina's rap music recommendations, despite being the most numerous, is the second lowest among all of her recommended genres and has a very small variance. This means that Lina's danceable album recommendations will predominantly and confidently be rap, followed by pop/r&b, then the other genres. Anand and Andres were both recommended danceable electronic albums the most. Though Andres was recommended slightly more electronic albums, with a slightly higher average distance of his albums. A key difference between Anand and Andres' danceable album recommendations is that Andres' music taste leans more towards rap than Anand's, as it is his second most recommended genre. However, Andres' rap recommendations were not the highest ranked since that genre has the largest average distance for him. Following electronic albums, Anand's recommendations have relatively even spread across the other genres with roughly equal average distances. This could be indicative of an eclectic music taste.

In the high instrumentalness results, Lina's recommendations show a much more eclectic music taste since her recommendations include 4 genres that Anand's and Andres' do not: pop/r&b, rap, global, and folk/country. It was only metal albums that she was not recommended. Additionally, all of Lina's recommendations have similar average distance and variance, indicating that her instrumental music taste spans evenly across genres. Anand and Andres have a lot of genre overlap, with Anand being recommended slightly more metal, experimental, and rock albums, and Andres being recommended slightly more electronic albums.




### ***Does our music taste better align with critics for certain genres?***

To answer this question, we examine how the album recommendation genre varies with average score. For danceable albums, we see that Lina's rap, no genre, and global album recommendations had the highest average score. This indicates that when it comes to danceable music, Lina has better taste when it comes to these genres. Anand's taste in electronic, rock, no genre, global, and experimental music all have higher average scores relative to other genres and other users, indicating these genres are better aligned with Pitchfork critics. By similar analysis, Andres' critically aligned genres include electronic, rock, and no genre.

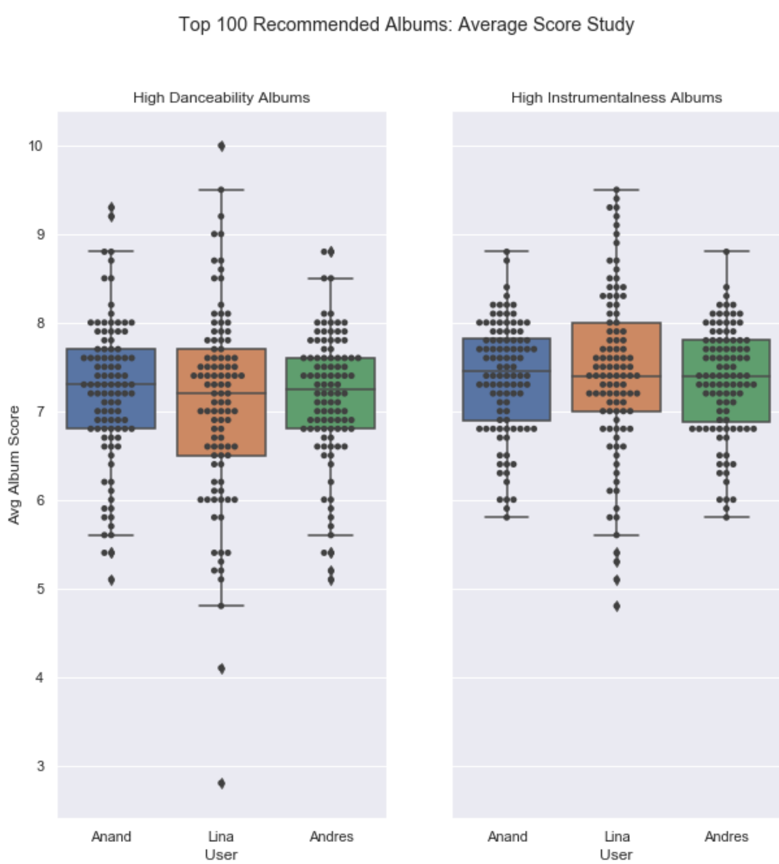



For instrumental album recommendations, Lina's music taste aligns most heavily with critics when it comes to experimental (in which her average score was over 8, usually Pitchfork's unofficial threshold for best new music), jazz, and global music. Anand and Andres' average score by genre almost entirely overlap, indicating that their music taste in instrumental albums is relatively similar when compared to that of critics.

## Score Analysis

### ***Whose recommendations have the highest scores?***

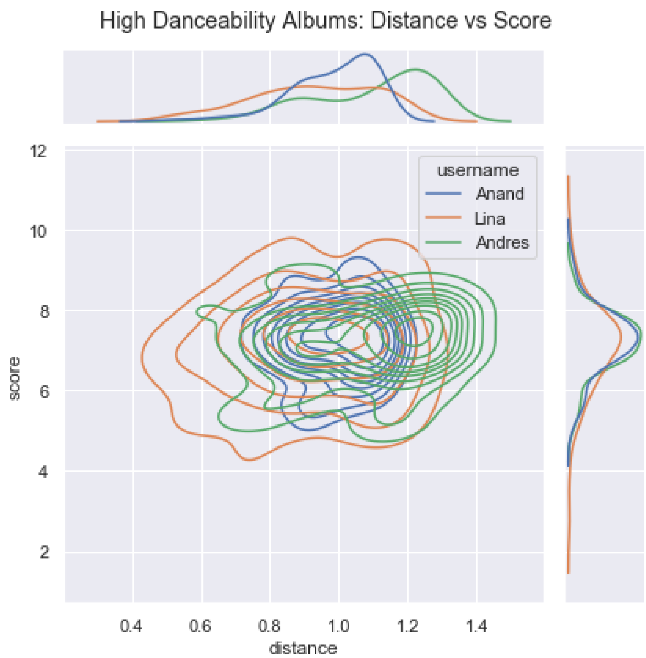
To see whose music taste best aligns with Pitchfork critics, we examine the average score of each user's recommendation in the boxplot "Top 100 Recommended Albums: Average Score Study" for danceable and instrumental recommendations. For danceable albums, the order of the highest average mean score was Anand, Andres, and Lina. Although Lina had higher scores in her top 75% quartile albums, her music taste also included very low scoring danceable albums and more varying low album scores concentrated further away from good scores around her mean. Andres' scores remain clustered around his mean score with few outliers for danceable albums and no outliers for instrumental albums.

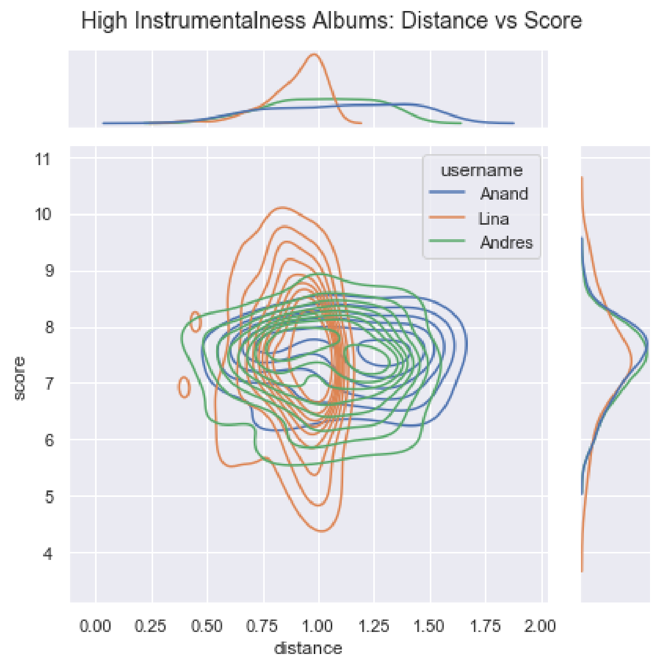



In the instrumental albums, we can see great overlap between Anand and Andres' album scores. This might indicate that their recommendations will heavily overlap. The few differences between their albums involve a few higher scoring albums for Anand, pushing his mean score here to the highest. However, Lina's instrumental albums vary the most in terms of score, with many far exceeding Anand's average score, Overall, Lina's above average scoring instrumental albums best align with critics since their scores are high and often far exceed 8 (the unofficial score cutoff for best new music).

***Is album recommendation distance influenced by the album score?***

Is there a relationship between how highly an album is recommended to a user and album score? Are some users more strongly recommended critically acclaimed albums over others? The joint KDE plots below examine the danceability and instrumentalness album recommendations for the 3 users and indicate how the distance, a measure of how far away an album recommendation is from the user's library, varies with the album's score. In the danceable albums on the left, we can see that all 3 users had recommended albums' scores roughly concentrated around 7. Albums scoring 7 could be recommended at most of the distances observed. This can mean that for danceable albums, having better or worse recommendations did not have a relationship with album score. Higher scoring albums did not appear to be more or less recommended. In the instrumental albums recommendations, Lina's distances were heavily concentrated around a distance range 0.5 and 0.1 but her scores varied widely, from very good to very bad, for these albums. This does not tell us that album distance has a relationship with score for her data. Anand and Andres' instrumental plots follow a similar pattern to the danceability albums plots; not indicating any relationship between album distance and album score.



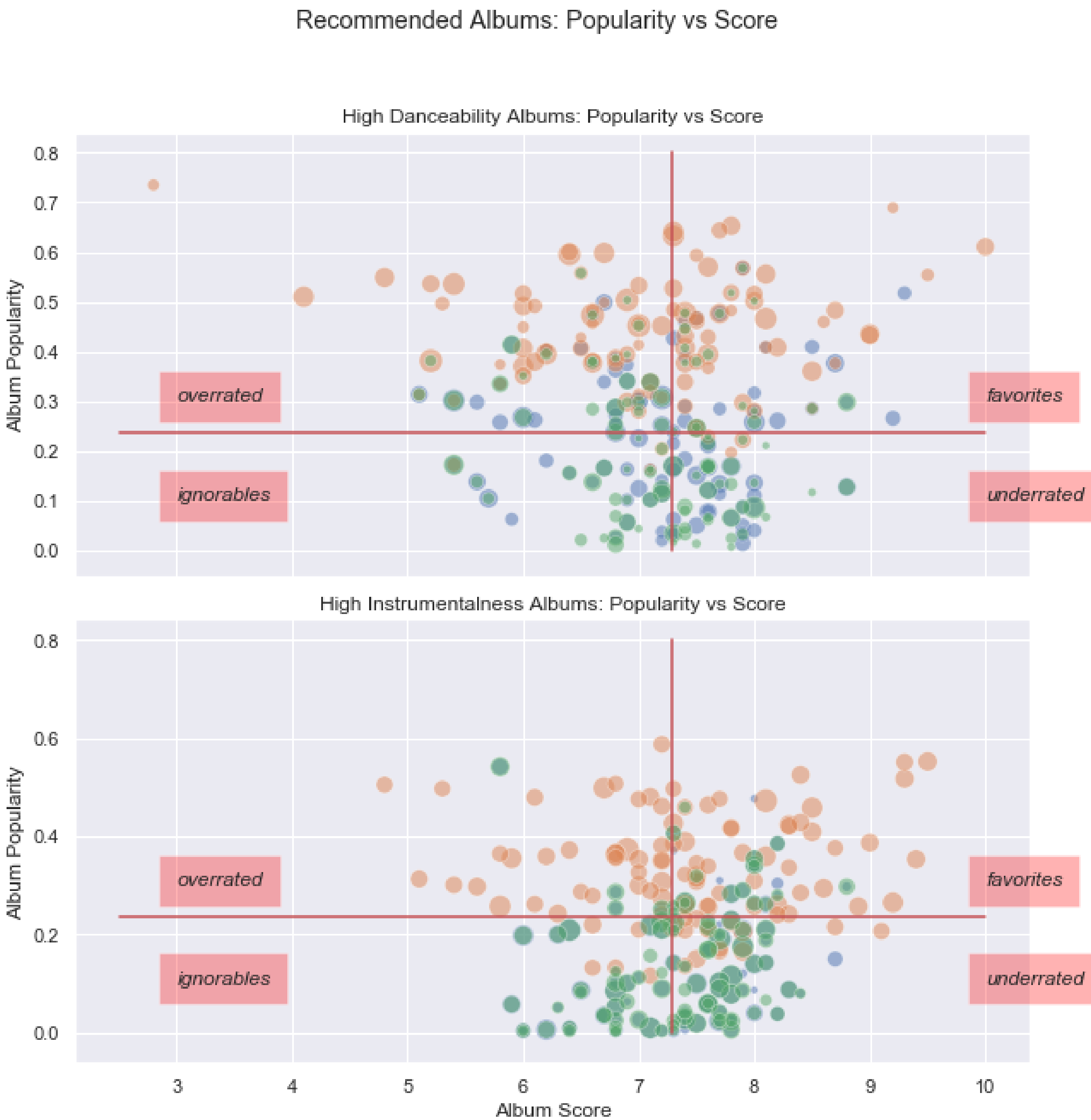

Group Label Analysis

***How would Pitchfork categorize our music taste  based on group_labels?***

Based on Pitchfork critics' album reviews in 2017, would critics classify the albums recommended to us, based on our music taste, as their "favorites", "underrated", "overrated", or "ignorable". The scatter plots below describe where each of our recommendations fall in terms of score and popularity. The red lines drawn indicate the average score and popularity for 2017 Pitchfork albums, and separate out our plots into the group labels. The plot contains markers for each user (Anand: Blue, Lina: Orange, Andres: Green) and the larger the marker the smaller the album's distance (i.e. it's more recommended to the user).

In the danceability recommendations plot, we can see that Anand's music taste represents each label fairly evenly but he has more underrated albums than others and the distances of those recommendations seems to be greater. This indicates that underrated album recommendations are slightly more aligned with his music taste. His other labels appear in the following order: overrated, favorites, ignorables. Anand's recommendations seem to indicate that when he listened to unpopular music, it is more critically acclaimed but his popular music tends to be slightly more overrated than critically acclaimed. Lina's danceable albums concentrate very heavily on either overrated or favorites. This supports the idea that her music taste emphasizes popularity more, and can span from critically acclaimed or critically panned. Andres' music taste is similar to Anand's; he also has underrated albums mostly. However, Andres' listens to slightly more underrated and ignorable music than Anand and less favorites than Anand. This means that Andres' danceable albums music taste does not emphasize popularity as much.

In the instrumental albums data, Lina's music taste follows the same trends towards popular music, whether it be overrated or favorites. She might, however, enjoy more underrated instrumental albums than underrated danceable albums. Anand's and Andres's instrumental album preferences almost entirely overlap. Again, Anand slightly prefers more favorite albums than Andres and Andres slightly prefers more ignorable albums than Anand. However, for instrumental albums, both of these users heavily prefer unpopular music (underrated, then ignorable).




Note that the appendix (Fig C.1)contains a bar chart for these group labels to assess the numbers.

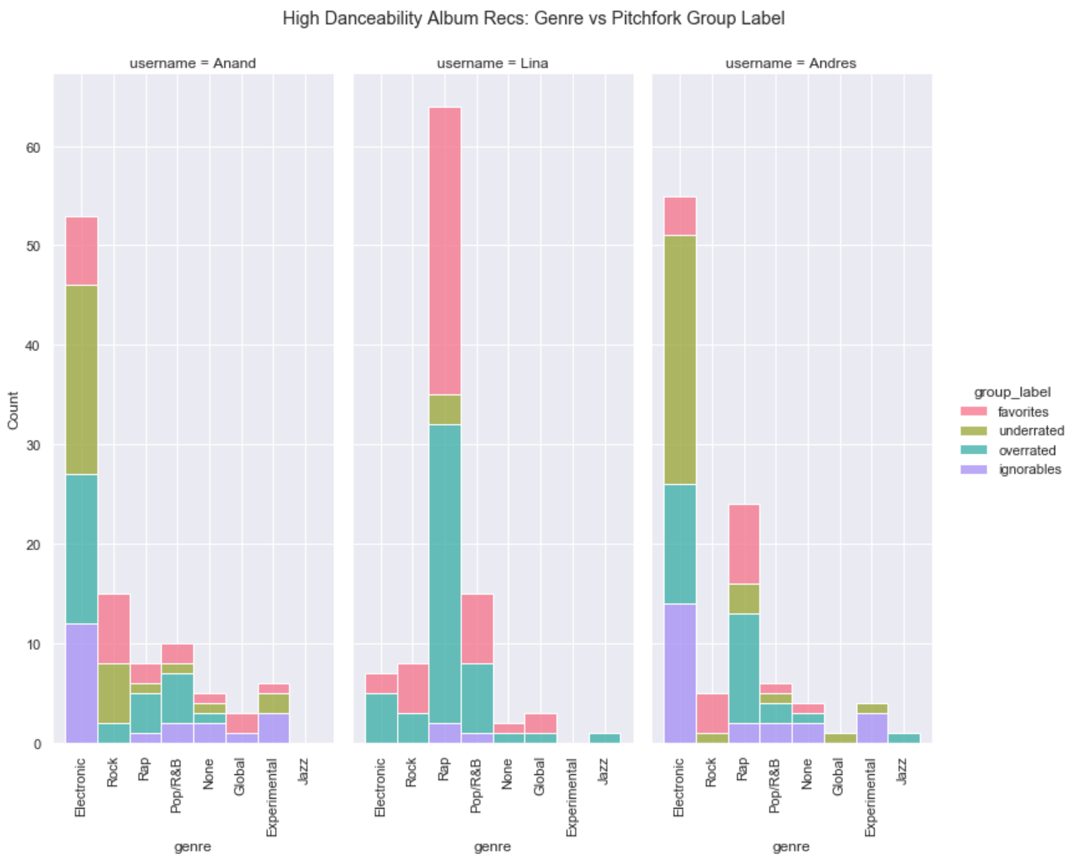
***Do group labels differ based on genre?***

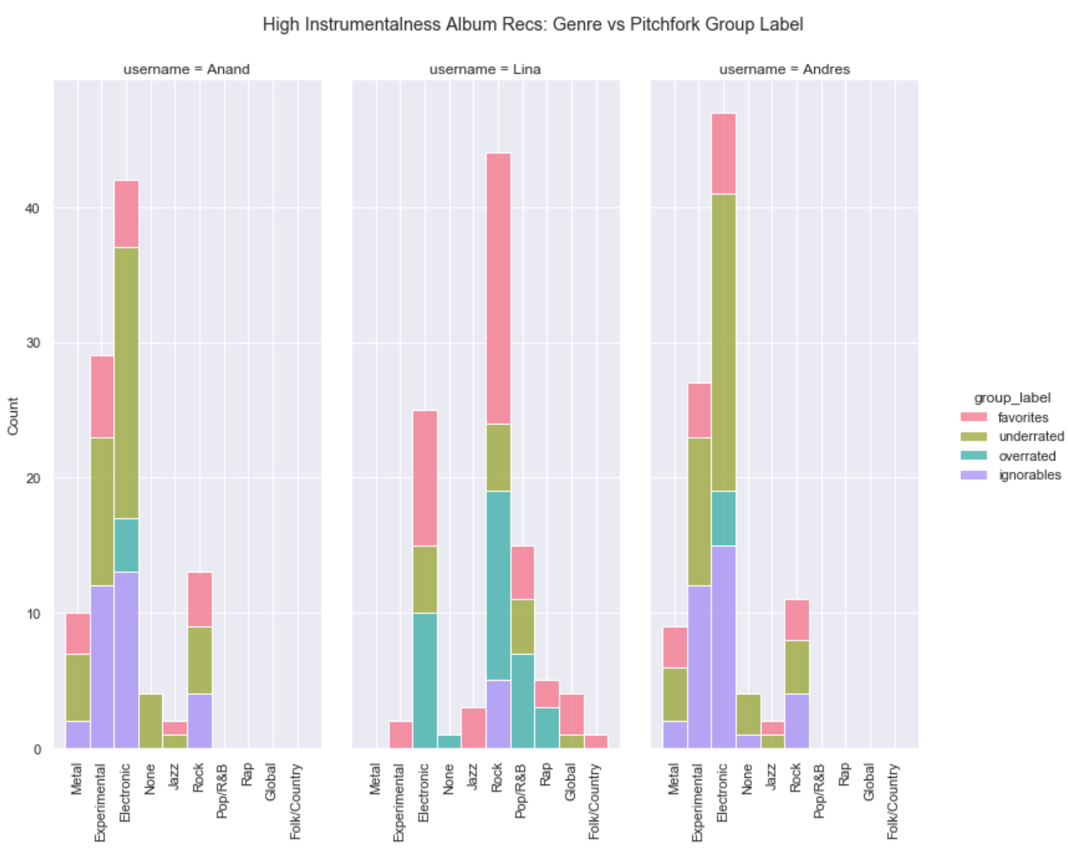
To determine whether each user's group labels change based on genre, we can observe the stacked bar plots of genre vs group label on the next page.

In the danceable album recommendations, Anand's most represented genre of electronic music has roughly even group labels, with favorites being slightly less represented. In his rock music recommendations, he was mostly recommended favorites or underrated albums; which is more closely aligned with Pitchfork critic scores. In the other genres, there was a predominant presence of overrated albums. The only exception was experimental music, where the recommendations were either critically acclaimed (favorites or underrated) or ignorable by critics and the public alike. Lina's results show an overwhelming and equal preference for favorites and overrated music. This trend is visible in her most prevalent genre, rap, but extends to every genre recommended to her except jazz, her least represented genre for danceability. Andres's label distribution roughly matches Anands in electronic music, their most recommended danceable genre. Andres's danceable rock music also matches Pitchfork critics since it is mostly labeled as favorites then underrated. However, in Andres' rap album recommendations, his second most represented genre, his preferences lean heavily towards favorites or overrated albums. This shows that when it comes to danceable rap albums, Andres's recommendations tend to emphasize popularity.

In the instrumental album recommendations, Anand and Andres' group labels by genre are almost identical. This further supports them having heavy overlap in the albums recommended to them. The only significant differences stem from Anand being recommended slightly more experimental favorites and Andres being recommended slightly more ignorable electronic music. Lina's instrumental album recommendations are represented the most by rock, electronic, then pop/r&b. In all 3 of these genres, she is mostly recommended both favorites and overrated albums. This further supports the idea that her instrumental music taste leans more towards popularity. However, her instrumental album recommendations feature more underrated albums, present in all of her top 3 genres.

Between both danceable and instrumental album recommendations, it seems that there can be changes in group label between genres and between these types of albums. However, Lina's inclination towards popular music is present in most of her genres. At a summary level, Anand & Andres' inclinations towards underrated and ignorable albums, low popularity music, is predominant in both data sets.


High Danceability Album Recs: Genre vs Pitchfork Group Label
username = Anand
username = Lina
username = Andres
Count
genre
Electronic
Rock
Rap
Pop/R&B
None
Global
Experimental
Jazz
group_label
favorites
underrated
overrated
ignorables


High Instrumentalness Album Recs: Genre vs Pitchfork Group Label
username = Anand
username = Lina
username = Andres
Count
genre
Metal
Experimental
Electronic
None
Jazz
Rock
Pop/R&B
Rap
Global
Folk/Country
group_label
favorites
underrated
overrated
ignorables

Qualitative Word Clouds:

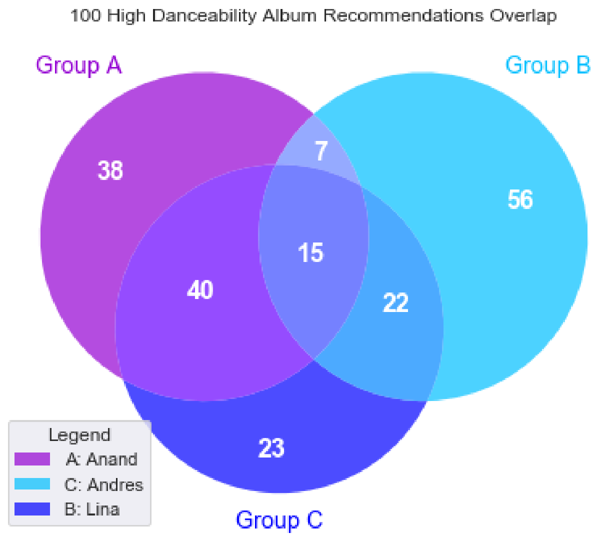

***Among our top 100 danceability recommendations, what do the reviews look like for our top 5 most critically acclaimed albums?***

Pitchfork features lengthy album reviews by the critic that justify the scores given to the album. For our danceable album recommendations, which had more differences between the 3 users than the instrumental album recommendations, do the reviews for our top 5 critically acclaimed recommendations share any similarities in words used to describe “good” danceable albums in the Pitchfork review?

Above we can see each user’s word cloud for their top 5 recommended danceable albums by score. This analysis is observational, so the conclusions drawn are qualitative. Both Lina’s and Andres’ best recommendations can be described with “rap”, which indicates that their best danceable albums drew from this genre heavily. Lina’s featured the names of some well known rap and pop/r&b artists like Beyonce and DMX, while Andres’ has more non-specific words on sub-genres like ‘techno” and other general terms. Anand’s word cloud differs from the others due to multiple genres being mentioned: “pop”, “rock”, “electronic”. There are some artists or albums mentioned (stereolab, mezzanine, buckingham). An observation that could be made is that Anand’s review word cloud seems to go in a different direction by genre than Lina’s and Andres’, and terms represented are dominated by music specific general terminology or genre specific terminology. Larger images of the word clouds are present in Appendix C.

## Do our specific album recommendations overlap?

***Danceability Recommendations***




Despite our danceability audio features being very different in our earlier analysis, it appears that sorting our Spotify libraries by high danceability, creating unique user library vectors to capture this subset of our libraries, and searching for similar albums in the Pitchfork 2017 reviews actually yields overlapping recommendations. Sorting on high danceability and returning 100 recommended albums each, out of 1009, actually results in almost 41% overlap in the unique albums returned. The returned album recommendations provide a good balance of personal album recommendations and albums that multiple users will enjoy.

In the overlap between Anand and Lina, the best album recommendations by score include Fleetwood Mac's *Tango in the Night* and Syd's *Fin*. The recommendations are mostly dominated by rap and pop/r&b albums. Between Anand and Andres, the best recommendations are Jlin's *Black Origami* and The Other People Place's *Lifestyles of the Laptop Cafe*, both are electronic albums classified as best new music. Between Lina and Andres, the best recommendations are 2 Chainz' *Pretty Girls Like Trap Music* and G Perico's *All Blue*, both higher scoring rap albums. Most of their overlap is around rap albums. Between all 3 users, the recommendation genres represented vary widely. Some top titles include *Doing It in Lagos: Boogie, Pop & Disco in 1980's Nigeria*, an disco album from Nigeria in the 80's, Mac DeMarco's *This Old Dog*, Yaeji's *Yaeji EP*, and Abra's *Rose*.

***Instrumentalness Recommendations***

In appendix C, we have our album venn diagram for the instrumentalness album overlap (see Fig C.5: Album Recommendation Overlap for top 100 High Instrumentalness Albums.). The results of the diagram confirm the theory that Anand's and Andres' instrumental albums will overlap greatly.

With 90 recommendations in common, the overlap for Anand and Andres point to some shortcomings in the album recommendation system. 1) The number of albums to pick from on the Pitchfork side is relatively small compared to everything available on Spotify, so reducing to 2017 makes this problem worse. 2) Album audio features are produced by averaging out the track audio features, which will collapse some of the uniqueness present. This system could likely give more unique recommendations for users if it searched the P4K 2017 album tracks for recommendations. That data is available, and it includes over 12,000 tracks for just 2017. 3) Audio features could likely be a limited method of capturing music taste, and struggles to capture nuance to lead to more unique album recommendations when simply searching for something with high instrumentals. It is likely that the overlap could stem from their instrumentalness distributions being very similar.

For the instrumentalness recommendations, Lina and Andres only had 1 album overlap: Bicep's *Bicep*. Recalling from our library analysis earlier, we all had this artist present in our spotify libraries. However, only Lina's and Andres' instrumentalness library search vectors ended up capturing their album's audio feature. This points out the 2nd shortcoming because the track characteristics in our library might be lost in the aggregation of P4K album audio features by tracks, and it points out the 3rd shortcoming since audio features might not be capable of returning recommendations that are already in our library, which has happened for other 2017 albums present that we have multiple songs from in my library.

## CONCLUSION

We can conclude that our music preferences are different or similar depending on how we measure our music features. There are certain music qualities that we converge on such as energy, tempo, and speechiness.  While we diverged on other qualities such as instrumentalness.

For the Pitchfork album recommendations, we found that albums returned off of a high danceability search yielded a good mixture of unique recommendations between the users and insightful overlap on albums multiple users would enjoy. The high instrumentalness recommendation results yielded 90/100 albums that both Anand and Andres would enjoy, which is unlikely. This could stem from the limitations of the recommendation system.

## REFERENCES


1. Our GitHub Repo
2. Spotify Developer Website
3. Kaggle.com

# APPENDIX

## Appendix C: P4K Analysis

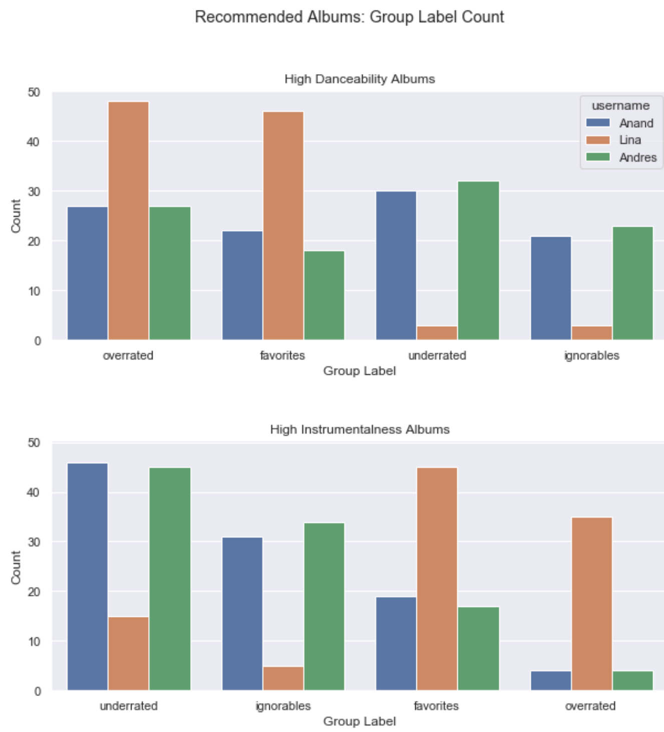



Fig C.1: Group Label count per User among 100 album recommendations, High Danceability and High Instrumentalness Album recommendations shown top and bottom respectively.

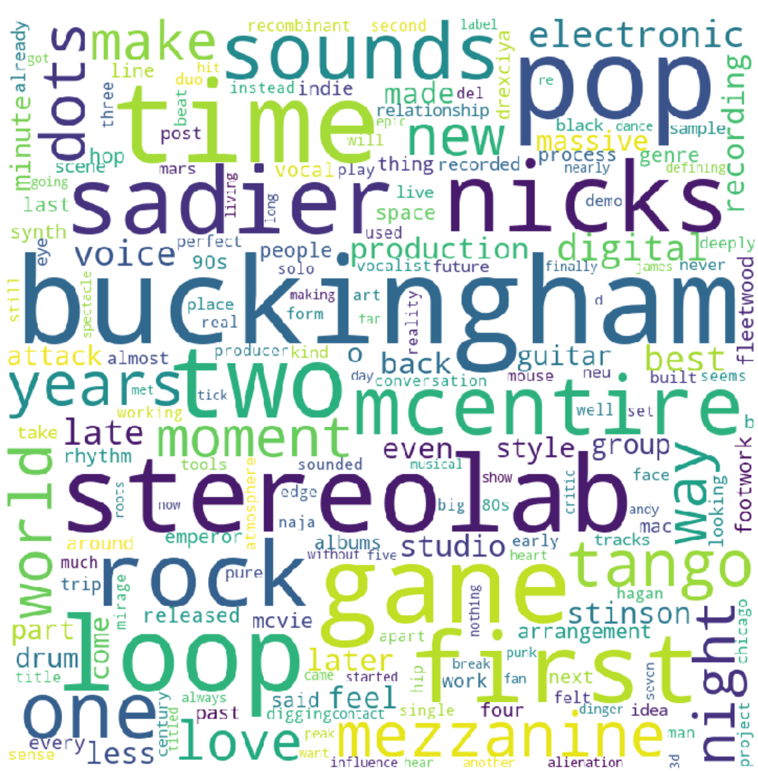


Fig C.2: Word Cloud for Anand’s top 5 album reviews from 100 High Danceability Album Recommendations


Fig C.3: Word Cloud for Lina’s top 5 album reviews from 100 High Danceability Album Recommendations

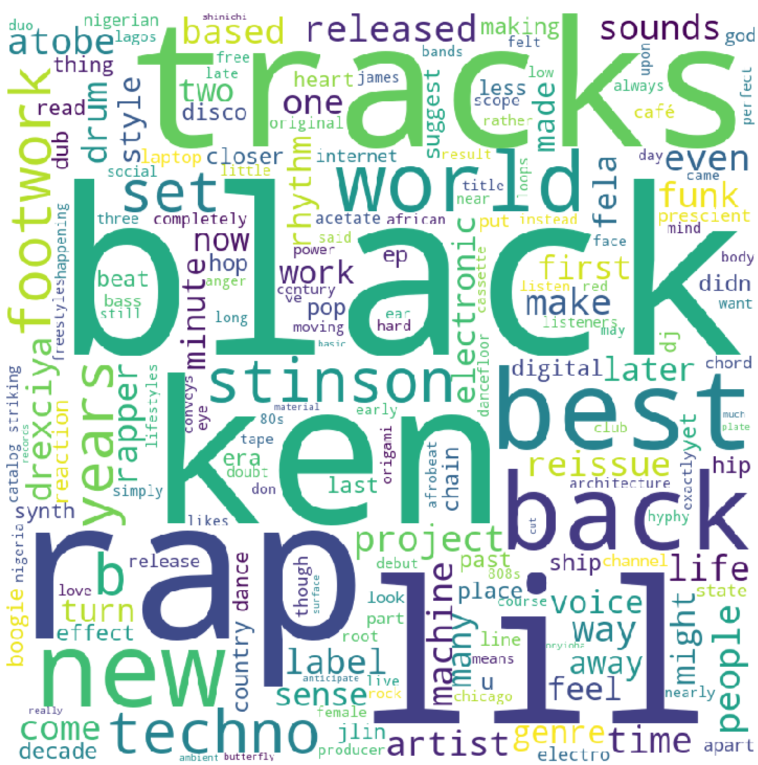


Fig C.4: Word Cloud for Andres' top 5 album reviews from 100 High Danceability Album Recommendations

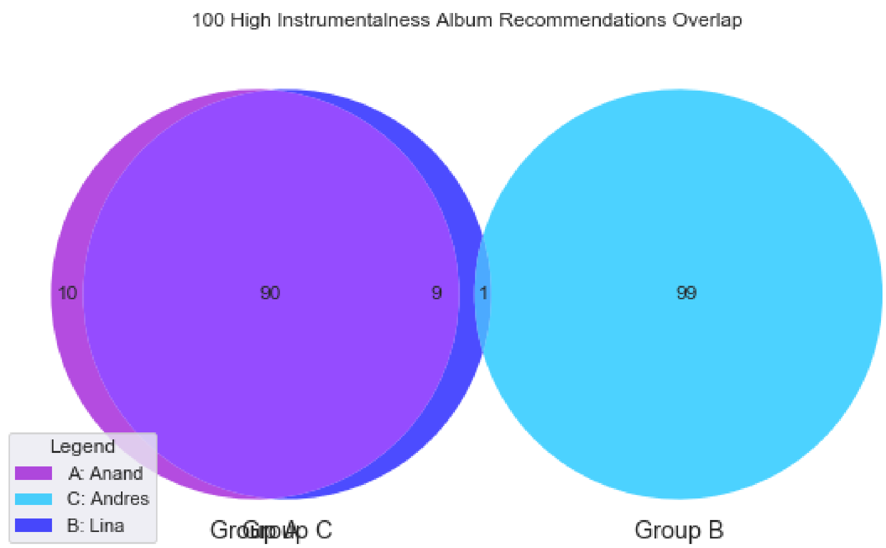



Fig C.5: Album Recommendation Overlap for top 100 High Instrumentalness Albums.