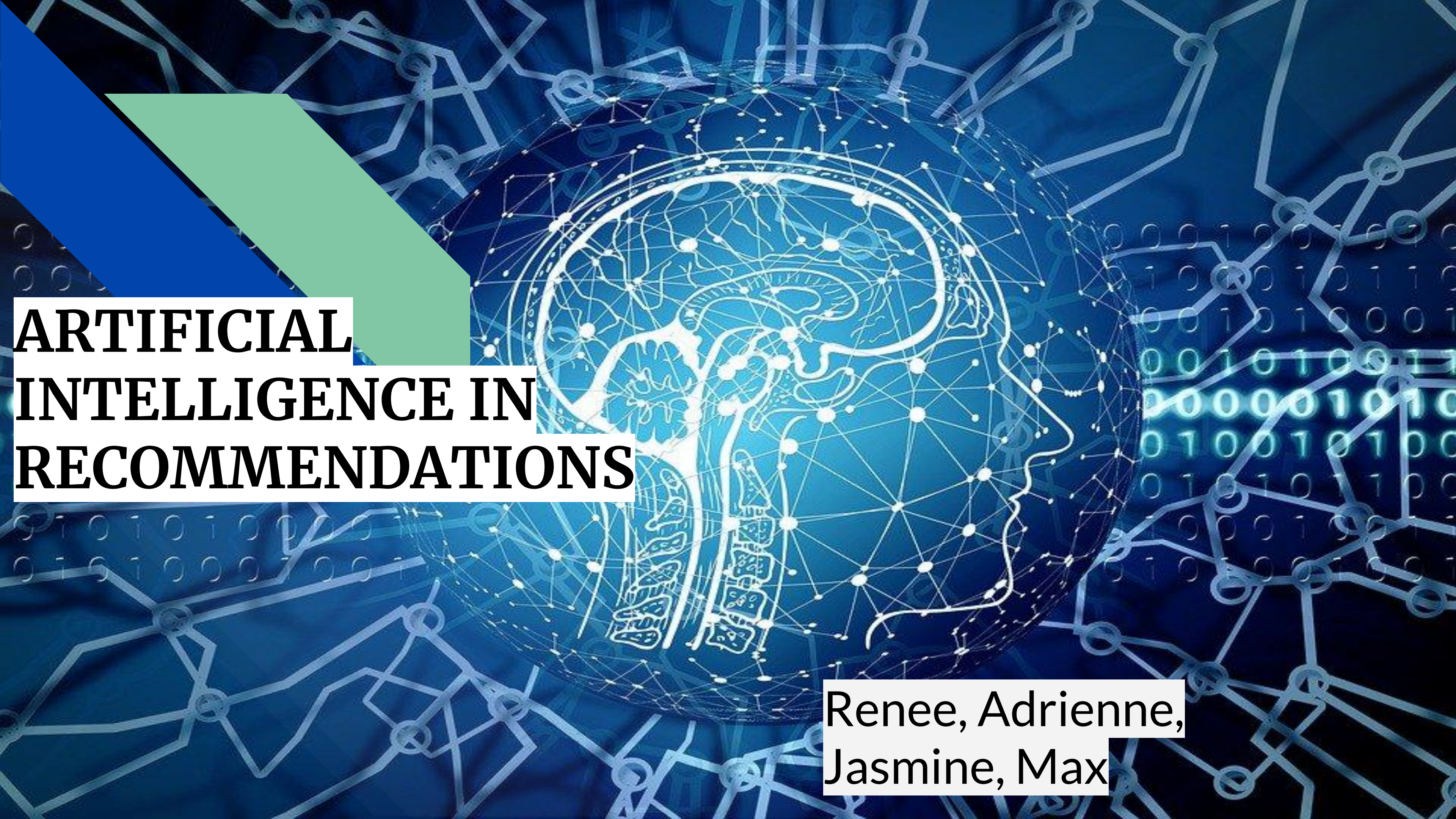


ARTIFICIAL INTELLIGENCE IN RECOMMENDATIONS

Renee, Adrienne, Jasmine, Max

The Issue at Hand

1. The music industry has a habit of only featuring artists that are marketable -
   a. Marketable traits are typically: straight, cis-het, white male artists
   b. These underrepresented musicians contribute great work to the musical scene but are routinely overlooked by mainstream media
2. Considering data collected from our peers, we found out what is desirable in a streaming service
3. We analyze the data to create our own music recommendation service would recommend underrepresented artists, but make music of the same caliber

Underrepresented: Women in Music

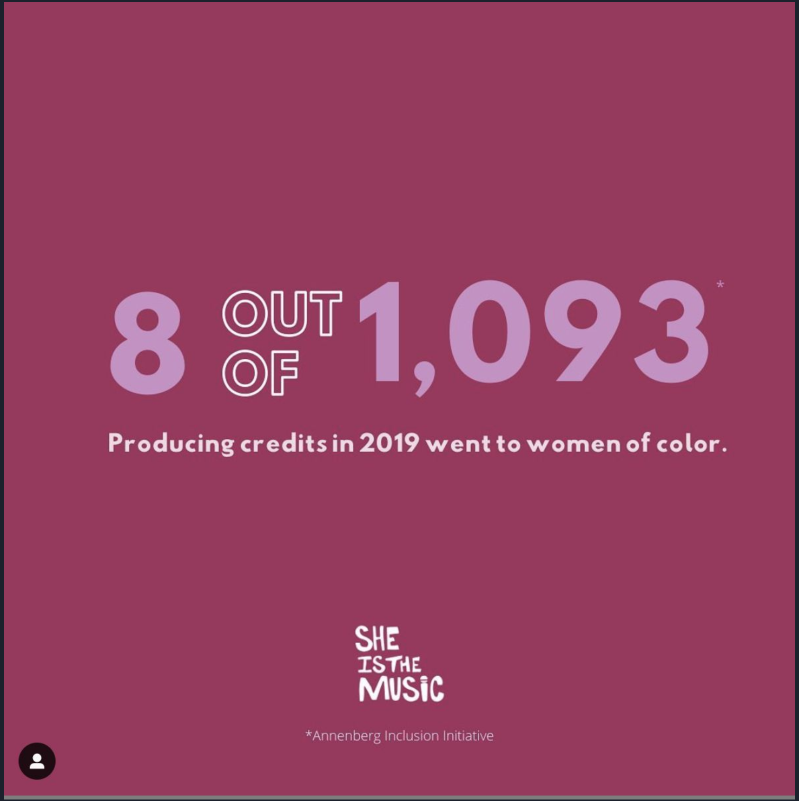
Women are routinely discriminated against in music

Despite the prominence of female artists, statistics actually show that:

- women are less able to pursue a successful musical career
- Men are favored in producing, record signing, and collaboration with other artists.
- Even when corrected for representation, men release more music
- Annenberg Institute found that, concerning gender women were more likely to appear as soloists than in a group.
- 12.3% of songwriters were female in the study, 43% of those were women of color.

Underrepresented: women in music distribution

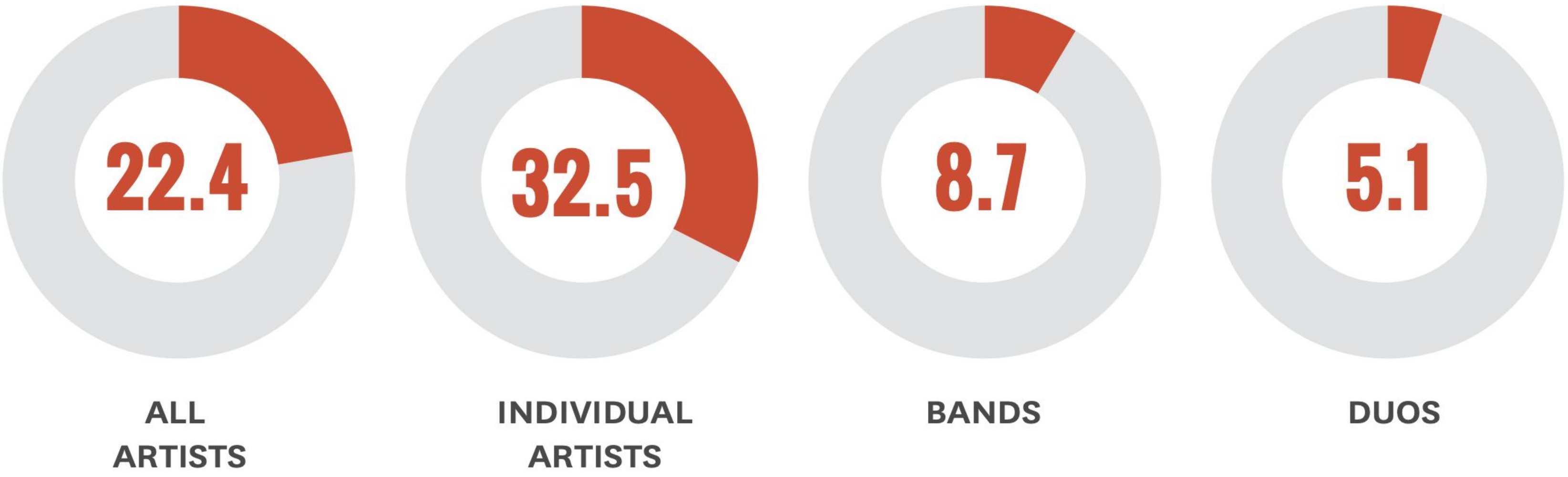
- Men are more likely to be able to exhibit “star behavior” like producing music, collaborating with other artists, and signing to record labels
- One explanation for the lower output of music made by females is that out of 5000 record labels, only one third had (ever) signed at least one female artist.
- All of these disparities keep women on the periphery of the music scene, and these factors are created, exacerbated, and perpetuated by the systems of inequality and bias that bar women from advancement in all careers



Point 1: underrepresentation

FOR FEMALES, MUSIC IS A SOLO ACTIVITY

Across 600 songs, percentage of females out of...



© 2018 DR. STACY L. SMITH

USC Annenberg
Inclusion Initiative

Why defining the gender of music matters

- Highlighting these differences between male and female artists can play a key role in the music industry.
- It can also solve the pay gap in the music industry.
  - Because music is an industry, there are discrepancies in how nonwhite, non-male people are paid for their work
- Changes in this industry could set high-visibility examples and facilitate transformation in other areas, too

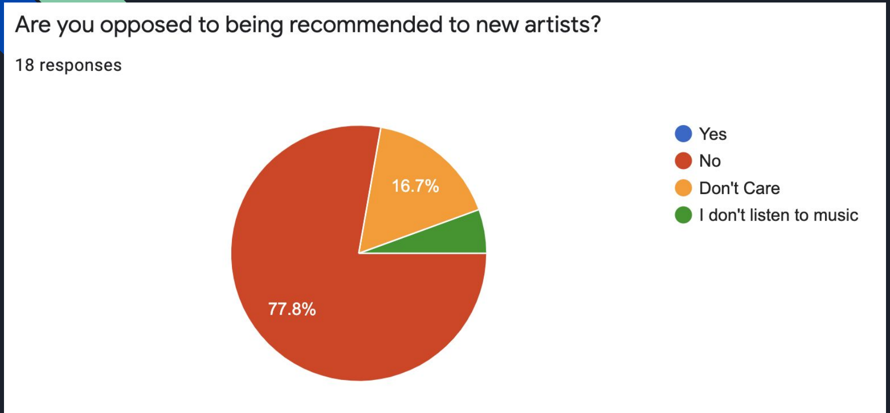
Point 1: Underrepresentation

- "What the experiences of women reveal is that the biggest barrier they face is the way the music industry thinks about women, the perception of women is highly stereotypical, sexualized and without skill. Until those core beliefs are altered, women will continue to face a roadblock as they navigate their careers."-Professor Stacy L Smith

- "Ideally, someday we'll define a 'female way' of producing music that could help women advance better in this industry and support a broader range of talent in the global music scene." - Agnes Horvat, Assistant Professor of communications at Northwestern University

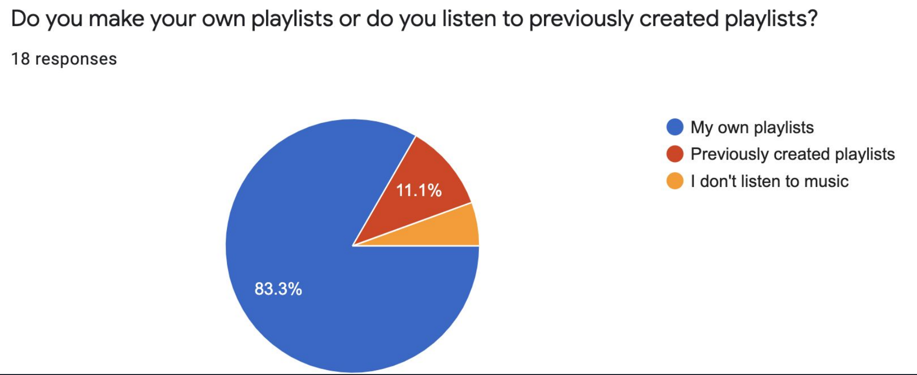
- "Once again this year we see a lack of female voices in popular music, however, one positive finding is that of the female performers in 2018, 73 percent were women of color. This seven-year high point reveals that the music industry is including women of color in ways that other forms of entertainment are not."
-Professor Stacy L. Smith

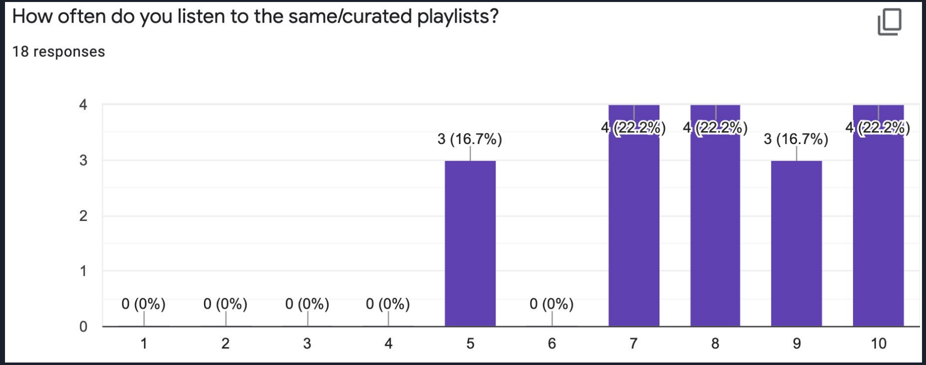
We polled the class and discovered that:



Most pollers are open or eager to be recommended new music - so a music recommendation service would be a good idea

Voters have to curate their own playlists, which proves to be somewhat of a downside of the platforms from which they stream music, for it limits their access to music





Streaming services

Most voters used Spotify as their primary music platform, while some used Apple Music, Pandora, and Sirius XM.

Their main issue with the services was that they could not recommend music, an inability to listen to specific songs but being forced to listen to playlists, and the dreaded advertisements that happen every ten minutes though listeners were guaranteed an ad-free thirty minutes

The best parts of the services were the wide library and availability of music. The platform’s interface allows for great interaction with the music listeners enjoy and the app recommends its own custom playlists to the listener as well

Our Data

	genre	artist_name	track_name	track_id	popularity	acousticness	danceability	duration_ms	energy	instrumentalness	key	liveness	loudness
110071	Pop	David Guetta	This Ain't Techno	26VQvcUzPMLc1Jd8oDEbfl	51	0.000510	0.592	219253	0.946	0.397000	D	0.329	-4.863
110443	Pop	T-Pain	1UP (feat. Profit Dinero)	2tljQIUKje5ni1pA147Ot0	56	0.023600	0.705	273810	0.584	0.000000	C	0.176	-6.397
110504	Pop	T-Pain	U Up	6p8JTP4A9NZHtxRVhkEo6s	55	0.007600	0.647	157390	0.484	0.000000	A	0.438	-10.316
110524	Pop	T-Pain	All I Want (feat. Flipp Dinero)	4FjcZsKyGhhZnuYq0nzXpZ	55	0.186000	0.514	161239	0.613	0.000000	E	0.284	-6.506
110588	Pop	T-Pain	It's My Dog Birthday	6Uh6VBPzqT8g4ADAYVpxeM	53	0.000582	0.604	211861	0.577	0.000000	G#	0.102	-7.647
...	...	...	...	...	...	...	...	...	...	...	...	...	...
152237	Pop	A Day To Remember	Sometimes You're the Hammer, Sometimes You're ...	6ZjvObA6zPCLnKvI0XYdcf	55	0.001310	0.194	274034	0.978	0.000000	C	0.369	-4.802
152252	Pop	Sonta	Crazy over You	2SolvGDc4dkRiii5bxGSkJ	56	0.559000	0.606	256675	0.578	0.000002	G#	0.153	-5.076
152256	Pop	Troye Sivan	LOST BOY	3KV9J5y3HDHKTOtjJtHwqi	54	0.047700	0.560	223794	0.889	0.000063	B	0.268	-5.148
152259	Pop	Tank	F***in Wit Me	3v10vlZlZPApvDz3kE4aNe	55	0.298000	0.364	242307	0.537	0.000000	F#	0.191	-5.912
152260	Pop	Kelly Clarkson	Whole Lotta Woman	1nukLnD50ey3rfs6jNnMJa	56	0.036900	0.838	173387	0.683	0.000000	A#	0.638	-6.205

mode	speechiness	tempo	time_signature	valence	Acousticness Difference	Danceability Difference	Energy Difference	Instrumentalness Difference	Liveness Difference	Loudness Difference	Tempo Difference	Valence Difference

Preferences Function

```
def classify(prefs): # preferences
  global songs

  songs = songsList

  songs = songs[songs["genre"] == prefs[0]]

  # popularity would be lower to let smaller artists be recommended
  score = np.mean(songs['popularity']) - 10
  print(score)

  songs = songs[songs["popularity"] <= score]

  songs = songs[songs['popularity'] >= (score - 20)]

  songs['Acousticness Difference'] = abs(prefs[1] - songs.acousticness)

  songs['Danceability Difference'] = abs(prefs[2] - songs.danceability)

  songs['Energy Difference'] = abs(prefs[3] - songs.energy)

  songs['Instrumentalness Difference'] = abs(prefs[4] - songs.instrumentalness)

  songs['Liveness Difference'] = abs(prefs[5] - songs.liveness)

  songs['Loudness Difference'] = abs(prefs[6] - songs.loudness)

  songs['Tempo Difference'] = abs(prefs[7] - songs.tempo)

  songs['Valence Difference'] = abs(prefs[3] - songs.valence)
```

```
songs = songs[~songs['track_name'].str.contains("Remix")]
songs['track_name'] = songs['track_name'].str.replace('$', 'S')
songs['artist_name'] = songs['artist_name'].str.replace('$', 'S')
```

```
def norm():
  global songs
  features = ['Acousticness Difference', 'Danceability Difference', 'Energy Diffe
  for i in features:
    songs[i] = (songs[i] - min(songs[i])) / (max(songs[i]) - min(songs[i]))
```

Get Possibilities Function

```
def getPossibilities(songName, artist_name, gen):
  song = songsList[songsList["track_name"] == songName]
  song = song[song['artist_name'] == artist_name]
  song = song[song["genre"] == gen]
  song.pop("artist_name")
  song.pop("track_name")
  song.pop("track_id")
  song.pop("popularity")
  song.pop("duration_ms")
  song.pop("key")
  song.pop("mode")
  song.pop("speechiness")
  song.pop("time_signature")

  person2 = []
  features = song.columns.values
  for i in features:
    person2.append(song[i].values[0])
  classify(person2)
  norm()

  return person2

getPossibilities("Californication", "Red Hot Chili Peppers", "Rock")

songs
```

Final Weights

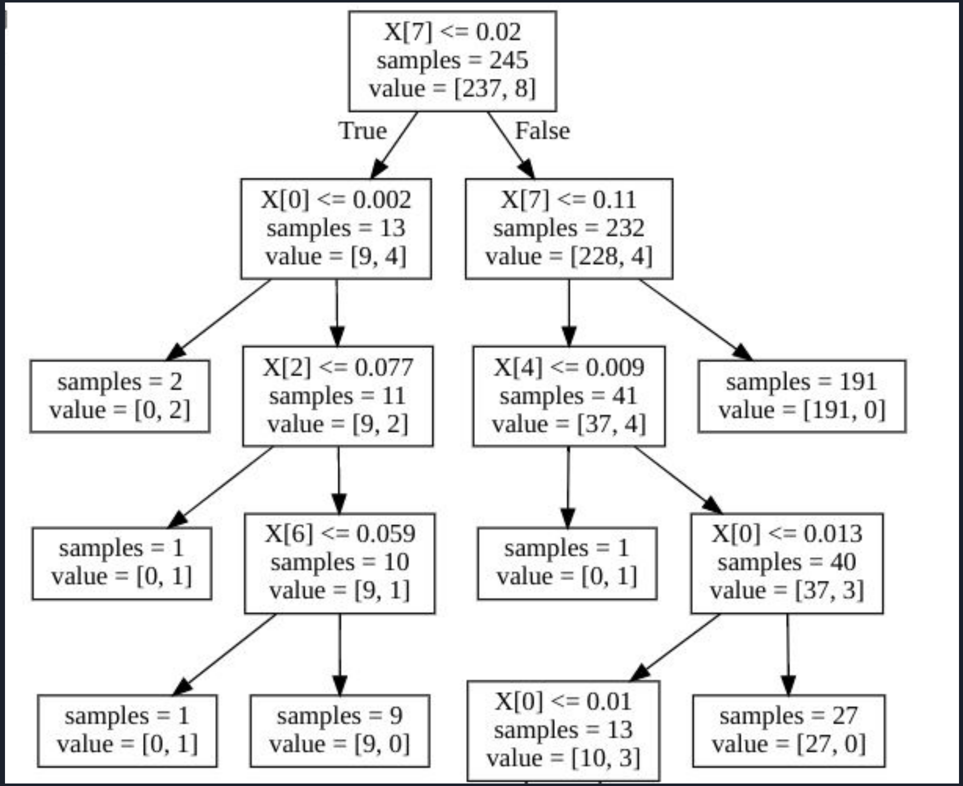
```
from sklearn.model_selection import train_test_split

# percentages based on 100% / 1.0
acoW = .14
danW = .16
eneW = .11
insW = .2
livW = .05
louW = .07
temW = .09
valW = .18

temp = songs[['Acousticness Difference', 'Danceability Difference', 'Energy Difference', 'Instrumentalness Difference', 'Liveness Difference', 'Loudness Difference',

X = temp.to_numpy()

songs['Final Weighting'] = (songs['Acousticness Difference']*acoW + songs['Danceability Difference']*danW + songs['Energy Difference']*eneW + songs['Instrumentalness

sort = songs.sort_values('Final Weighting')
print(sort)
sorted = sort['Final Weighting'].to_numpy()

temp2 = songs['Final Weighting']

y = []

for i in temp2.to_numpy():
  if i < sorted[9]:
    y.append('Recommend')
  else:
    y.append("Don't Recommend")
```

Decision Tree Classifier

```
from sklearn.tree import DecisionTreeClassifier

classifier = DecisionTreeClassifier(max_depth=5)

X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.8, random_state=42)

classifier = classifier.fit(X_train, y_train)

y_predict = classifier.predict(X_test)

accuracy = classifier.score(X_test, y_test)
print(accuracy)
```

```
0.9661016949152542
```

Decision Tree Visual



Final Recommendations

Sample Song

```
getPossibilities("Californication", "Red Hot Chili Peppers", "Rock")
```

Accuracy of Decision Tree

```
0.9032258064516129
```

Song Recommendation

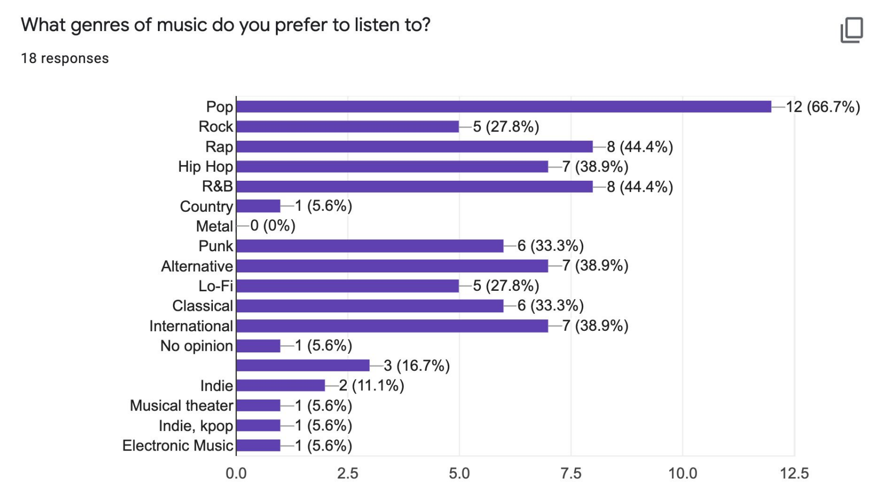
```
recLis = []
nonRecLis = []

for i in range(len(X_test)):
  if y_test[i] == 'Recommend':
    recLis.append(X_test[i])
  else:
    nonRecLis.append(X_test[i])

comp = songs['Acousticness Difference'].to_numpy()
track = songs['track_name'].to_numpy()
artist = songs['artist_name'].to_numpy()
for j in range(len(comp)):
  for i in recLis:
    if comp[j] == i[0]:
      print(track[j],artist[j])
```

```
Shine A Light Bryan Adams
```

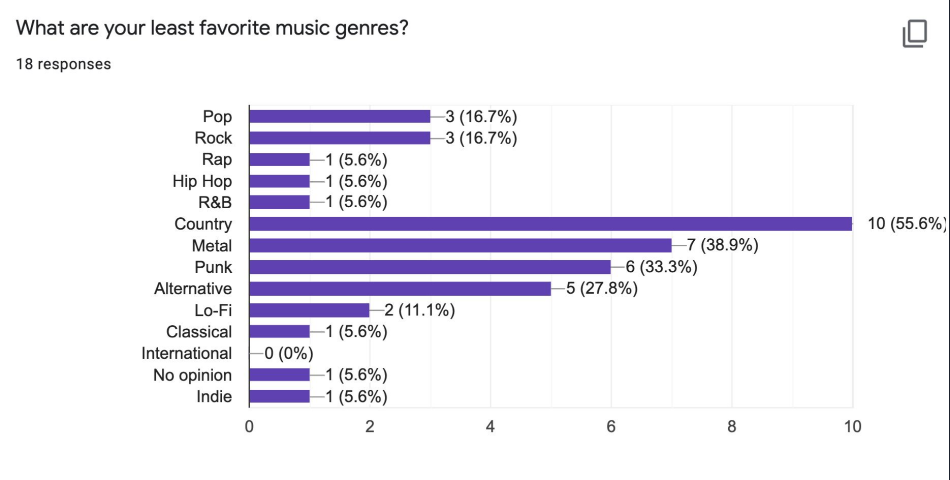
Successes and Failures

- The algorithm is able to output a recommendation that fits the general scope of the song inputted
- The accuracy varies depending on what song
- We couldn’t reach our goal of including more minorities in our recommendations because there weren’t features that contained gender or race, as an example
- We did try to include less popular artists with the popularity feature
- Couldn’t figure out a way for the algorithm to optimize the weighted averages on its own
- Songs that weren’t in the data set could not be used

We polled the class and discovered that:



The most popular music genre in AI4ALL 2020 was pop music, followed by rap, R&B, and Hip Hop



Country, metal, and punk were the least popular genres to listen to among the pollers

We discovered that:

Instruments and the background music/beats were the most important part of a song, followed closely by the lyrics:

- I believe that the instruments/beat is the most important in music. It'd be lyrics after that.
- instruments, beat, lyrics
- the melody,instrumentals, voice


Thank you

Sources

https://www.abetterbalance.org/equal-pay-day-2019-challenges-and-opportunities-for-progress/ Equal Pay for Equal work

http://assets.uscannenberg.org/docs/inclusion-in-the-recording-studio.pdf Anneberg Study

https://www.instagram.com/p/CBgs8Qmnfw5/ She Is the Music Instagram page

https://rachelcorbett.com.au/podschool-podcast-use-music-podcast/ musician image

https://www.kaggle.com/zaheenhamidani/ultimate-spotify-tracks-db Spotify Tracks Data Set