CS 209a Final Project: Spotify
Scope of Work & Preliminary EDA

Group Number

We are **Group 44**. Group members (alphabetical) are:

- Anna Davydova
- Michael S. Emanuel
- Avriel Epps
- Erin Williams

Group Communications

Our group has met in person twice. We have set up a team Slack channel for real time communications. We have established a GitHub organization for sharing source code, and created a repository for this project. We have also created a shared folder on Dropbox so the team can have synchronized access to large data files that are not suitable for GitHub. Our members live off campus and our numbers include three parents, so we've engineered our team workflows to allow us to collaborate efficiently when working from home.

TF Communications

Our group has been in contact with our TF, Will Claybaugh. We had an initial contact with him when we discovered that the link to the Spotify Million Playlist Dataset had expired because the underlying competition ended on June 30, 2018. Will kindly provided us a link to an archived copy of the data. We had a very insightful follow up meeting with Will on October 22, 2018 where we discussed the initial direction of our project. We plan to schedule another meeting over the next two weeks and to keep Will in the loop on our progress.

Project Question and Objective:

- **Project Question**: In this project we will develop and test a model that answers the following question: **Can Spotify improve end-of-playlist track recommendation diversity, without sacrificing user engagement?**
- **Motivation**: It is no secret that the majority of streaming playlists are dominated by popular tracks that hinder user discovery of new and less well-known artists. As The Guardian succinctly puts it when comparing music experience before the rise of streaming service and to now: "*What's different now is that the dominance of streaming rewards passivity – repeat listening – rather than active discovery*"[1]. Similarly, Spotify has publicly expressed concerns about the gender bias in their algorithmically curated playlists, noting that female artists are recommended to users at significantly lower rates

[1] "Ed Sheeran has 16 songs in the Top 20- and its a sign of how sick the charts are.", Laura Snapes, March 2017, https://www.theguardian.com/music/musicblog/2017/mar/10/ed-sheeran-has-16-songs-in-the-top-20-and-its-a-sign-of-how-sick-the-charts-are

[2] Garcia-Gathright, J., Springer, A., Cramer, H. (2018). "Assessing and Addressing Algorithmic Bias - But Before We Get There." *Computers and Society*.

than male artists [2]. The goal of this project is to explore different ways that Spotify can improve diversity of its recommendations while maintaining positive user experience and engagement.

- **Definitions of key metrics**: We define user engagement as uninterrupted playlist consumption with 30 seconds or more of per track listening time (this threshold is based on Spotify's economic model where 30 seconds "counts" as a monetized stream). In this analysis, we will measure diversity by artist tier level (as determined by popularity) and artist gender; however, we hope to build a model in which any measure of diversity (e.g., artist locale) can be plugged in and used as a metric for model success. We will set a specific percentage threshold after a more thorough evaluation of our data popularity rankings with a hope of creating 5 artist tiers that range from "mega-pop star" to "completely unknown".
- **Definition of success**: We will define success of our model in two ways. First, by comparing the diversity of track attributions defined above that our model produces in comparison to Spotify's current end-of-playlist recommendation system. Second, by comparing user satisfaction with the listening experience that our model produces as compared to the listening experience that Spotify's current model produces, as detailed below. If our model produces more diverse tracks and does not significantly negatively impact user satisfaction, we will consider our model a success.

Proposed methodology:

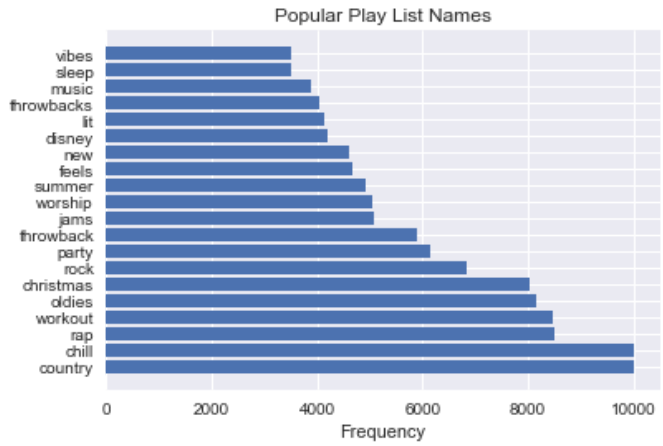
- **Part I: Build Simple Recommender**: The first part of our project will focus on building an accurate predictor model to recommend songs that the user already likes within the context of a playlist they are listening to. This will function similarly to Spotify's "Recommended Songs" feature which recommends songs to be added to every user-built playlist. Specifically, we would train the model on a portion of the user playlist and then test it by having it predict the missing tracks. We will score the predictions on the basis of their cross-entropy loss. We would also compute an "uninformed" model that had no information at all. This model would just guess the population average frequency for artists, albums, and tracks. We will then compute a dimensionless *entropy ratio* of the entropy of our predictions divided by the entropy of the uninformed model. A lower ratio would be better; zero would mean perfect accuracy on every prediction, one would mean no added information compared to the uninformed model. We will compute a blended loss function using the entropy loss ratio for the predictions of the artist, album and track. Proposed weights will be 30% for the artist, 20% for the album, and 50% for the track. We will score its accuracy not only on the first guess, but also allow the model to make up to 5 to 10 guesses (Spotify currently produces 10 recommendations at a time), and compute its accuracy on that basis. The rationale for this is that a recommender system that gave you 5-10 choices for the next track could provide a useful user experience. We will report these scores and highlight them in the model assessment we publish. Cross entropy loss is an excellent metric for model training, but end users (e.g. executives at Spotify) want to know the accuracy of these forecasts.

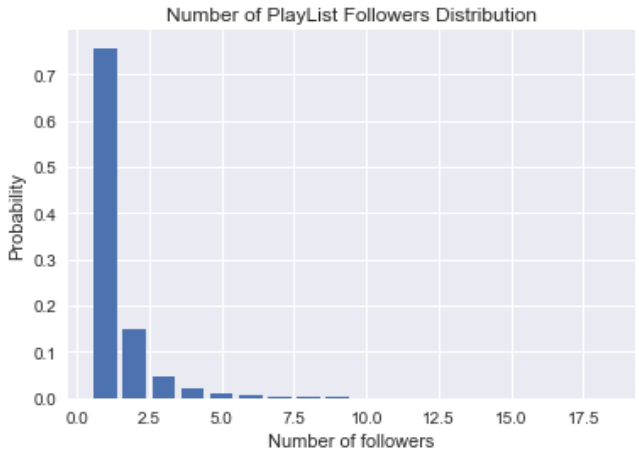
- **Part II: Incorporate Diversity:** In the second part of the project, the team will focus on introducing songs from the lesser known/popular artists and have recommender model produce a blended list of suggested tracks that includes tracks from the first part as well as less well-known tracks from our bounded “pools” of tracks. The boundaries we will set for such pool are tracks from mid-tier, female artists. If this approach is successful, a company like Spotify will be able to introduce any such set of artists based on their promotional themes and diversity goals.
- **Part III: Test Model Performance:** Since we are developing a brand new recommender system, we do not have a pre-existing test data set to validate our model’s performance. Thus, we plan to design an experiment with a randomly assigned sample of music listeners to test our product. Specifically, we will look to compare the user experience and engagement (how long and how many songs were consumed), with the non-diverse, simpler recommender from Part I vs. blended recommender Part II. We will devise a scoring mechanism that would allow us to get insights into how well/poorly our diverse recommender performs in comparison to what is currently already in the market (as captured by our model from Part I).

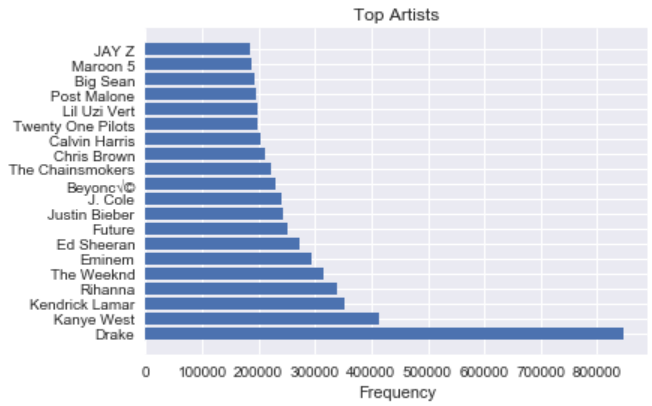
Data Snapshot: We have included several preliminary visualizations of our data. We are working to pull in additional data sets as discussed in the Action Plan below.

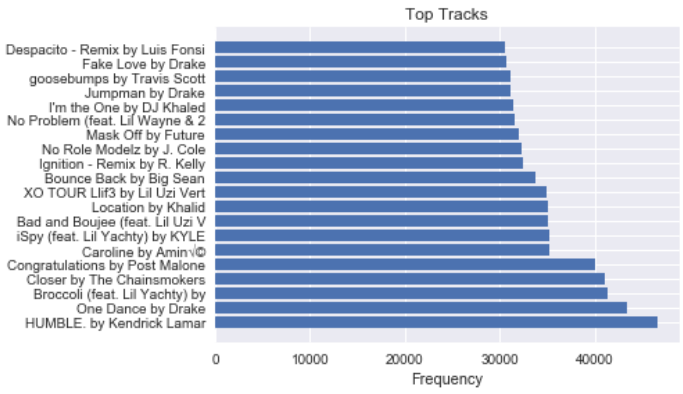
SummaryStats	
number of playlists	1.000000e+06
number of tracks	6.634643e+07
number of unique tracks	2.262292e+06
number of unique albums	7.346840e+05
number of unique artists	2.958600e+05
number of unique titles	9.294400e+04
number of playlists with descriptions	1.876000e+04
number of unique normalized titles	1.738100e+04
avg playlist length	6.634643e+01

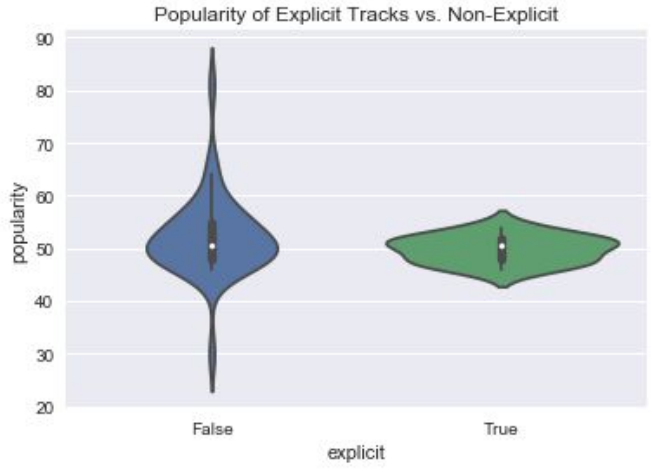
Our very early observations : 1) there is a larger spread in artist popularity vs. song popularity; 2) there are many track duplicates with number of unique tracks far below the total number of tracks. 3) there also appears to be many of the same titles or missing titles for different tracks as number of unique titles is lower than the number of unique tracks; 4) popularity of the track doesn’t seem to be significantly affected by whether or not it has explicit content unless you are in the tails (i.e. most popular and least popular tracks are clean); 5) Popularity of the artist appears to follow a normal distribution (albeit with leptokurtosis)

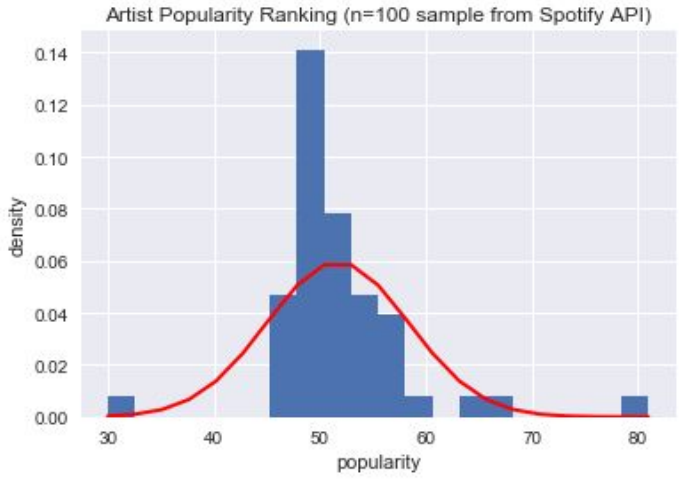

Popular Play List Names
vibes
sleep
music
throwbacks
lit
disney
new
feels
summer
worship
jams
throwback
party
rock
christmas
oldies
workout
rap
chill
country
0
2000
4000
6000
8000
10000
Frequency


Number of PlayList Followers Distribution
Probability
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
Number of followers


Top Artists
JAY Z
Maroon 5
Big Sean
Post Malone
Lil Uzi Vert
Twenty One Pilots
Calvin Harris
Chris Brown
The Chainsmokers
Beyonc√©
J. Cole
Justin Bieber
Future
Ed Sheeran
Eminem
The Weeknd
Rihanna
Kendrick Lamar
Kanye West
Drake
0
100000
200000
300000
400000
500000
600000
700000
800000
Frequency


Top Tracks
Despacito - Remix by Luis Fonsi
Fake Love by Drake
goosebumps by Travis Scott
Jumpman by Drake
I'm the One by DJ Khaled
No Problem (feat. Lil Wayne & 2
Mask Off by Future
No Role Modelz by J. Cole
Ignition - Remix by R. Kelly
Bounce Back by Big Sean
XO TOUR Llif3 by Lil Uzi Vert
Location by Khalid
Bad and Boujee (feat. Lil Uzi V
iSpy (feat. Lil Yachty) by KYLE
Caroline by Amin√©
Congratulations by Post Malone
Closer by The Chainsmokers
Broccoli (feat. Lil Yachty) by
One Dance by Drake
HUMBLE. by Kendrick Lamar
0
10000
20000
30000
40000
Frequency


Popularity of Explicit Tracks vs. Non-Explicit
popularity
20
30
40
50
60
70
80
90
False
True
explicit


Artist Popularity Ranking (n=100 sample from Spotify API)
density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
30
40
50
60
70
80
popularity

Action Plan for Milestone Three:

The tasks and deadlines below are preliminary and speculative. It is NOT expected that this is what will actually happen! This is rather intended to be the first iteration of a work plan that will evolve if necessary. We should assess at least once a week if we're on track with the plan, adjust it as necessary, and ensure that we're on track to product the Milestone 3 deliverables.

Data Collection (Oct. 25)

- Download Million Songs Dataset to team Dropbox. (**MSE, Oct. 16**)
- Download Million Playlist Dataset to team Dropbox.
    - Unpack the data. (**MSE, Oct. 18**)
    - Port Python utilities with the mpd to Python 3. (**MSE, Oct. 18**)
- Investigate the Spotify API; how do we access it? Which audio features are available? Do we want to download the audio features and ISRC numbers for our tracks?
    - Report; team decision on whether to download: (**AD, Oct. 23**)
    - Download data (**AD, Oct. 25**)
- Investigate Lyrics Wiki—can we link song lyrics using our available track data? (**AV, Oct. 25**)

Exploratory Data Analysis (Nov. 12)

- Cumulative frequency of genres, sub genres, artists, albums, and tracks: plot a curve showing the cumulative popularity of these three entities in descending order. Popularity measured by the sum of the number of followers on each track. Popularity of albums and artists is an aggregation of track popularity. Recommend a popularity threshold for us to limit the number of artists, albums, and tracks we consider in our models. (**EW, Oct. 28**)
- Scan data for missing/NA values and consider appropriate methods for dealing with missing values (i.e. remove outright, backfill, etc.) **(AE Oct.28)**
- Scan for duplicates (identical entries or same song listed either under individual artist or their band). **(AE, Oct. 28).**
- Additional Preliminary EDA on MPD (**EW, Oct. 28**):
    - Histograms of playlist length by number of tracks and in minutes. How long are these playlists?
    - Histogram of the number of followers. What is the distribution?
    - Explore most common track titles. How much do they cover? Can we glean useful information from the track titles?
    - Explore most common genres and their correlation with popularity?
    - Explore the purpose of the playlist (e.g. roadtrip, study, etc.) and their correlation with genres?
    - Create various stratified plots that overlay categorical features and explore interactions between variables.
    - Explore the distribution of types of artists on each playlist. What do they look like in terms of diversity (sub genre, artist tier, artist gender, etc).
- Preliminary assessment about Genre. Can we assign useful genre tags to artists, albums or tracks? How (e.g. ISRC numbers linking to another dataset)? Is it worth the effort? (**AE, Oct. 28**)

- Develop preliminary EDA strategy: Based on above work propose to rest of the team an EDA strategy, with additional EDA tasks as necessary. (**EW, Oct 30**).
- Carry out additional EDA tasks—assigned after above; completed by **Nov. 12**

***Baseline Model** (**Nov. 14**)*

- Python code to create and persist data frames for the following logical entities: artists, songs, tracks, playlists. Objective is to have two data frames for each type: a small one for quick exploration, and the full sized one. (**MSE Oct. 28**)
- Develop the uninformative model and cross entropy loss function; compute prediction accuracy for next 1, next 3, next 5. (**MSE Oct. 30**)
- First pass at predicting the artist, album, and track using "simple" techniques, e.g. a softmax classifier with linear weights of available features. (**MSE Nov. 12**)
- Review and critical feedback of Baseline Model; should be provided informally on an ongoing basis as code is posted to GitHub. More formal presentation of feedback, both an assessment on quality and suggestions about future work: (**AE and EW, Nov. 14**)
- Experiment with creating features via unsupervised learning (i.e. find the optimal number of k-means or hierarchical clusters for the playlists in the data set and use the resulting cluster assignment as an additional feature in our supervised learning model that follows).(**AD, Nov 12)**
- Design and implement schema for selecting tracks that would increase improve diversity of our model recommendations (**AE, Nov 12**).

***Milestone 3 Deliverables** (**Nov. 26**)*

- Description of Data for Milestone 3(**AE, Nov. 24**)
- Exploratory Data Analysis Findings for Milestone 3 (**AE, Nov. 24**)
- Preliminary Milestone 3 merging above and making one round of edits (**EW, Nov. 25**)
- Outline recruitment strategy and experiment protocol for final A/B testing (**AE, EW, AD, Nov 25**)
- Feedback / suggested changes before Milestone 3 deadline (**Everyone, Nov. 26**)
- Final version of Milestone 3 and submission on Canvas (**AD, Nov. 27**)