**"What we talk about
when we talk about successful playlist"**
-- Hong Wu (Interview Assignment)

DISCLAIMERS

With limitation/restriction from the given dataset, some estimation/projection are based open data source which may differ from actual Spotify cases. All charts, visualizations and conclusions are pure for demonstrating personal data analysis capabilities.

Assumptions on Data
With provided data dictionary, all data is on daily basis, projection made on time goes by 24 hours, 7 days a week , 31 days per month, simply for calculation convenience, open data is also used.

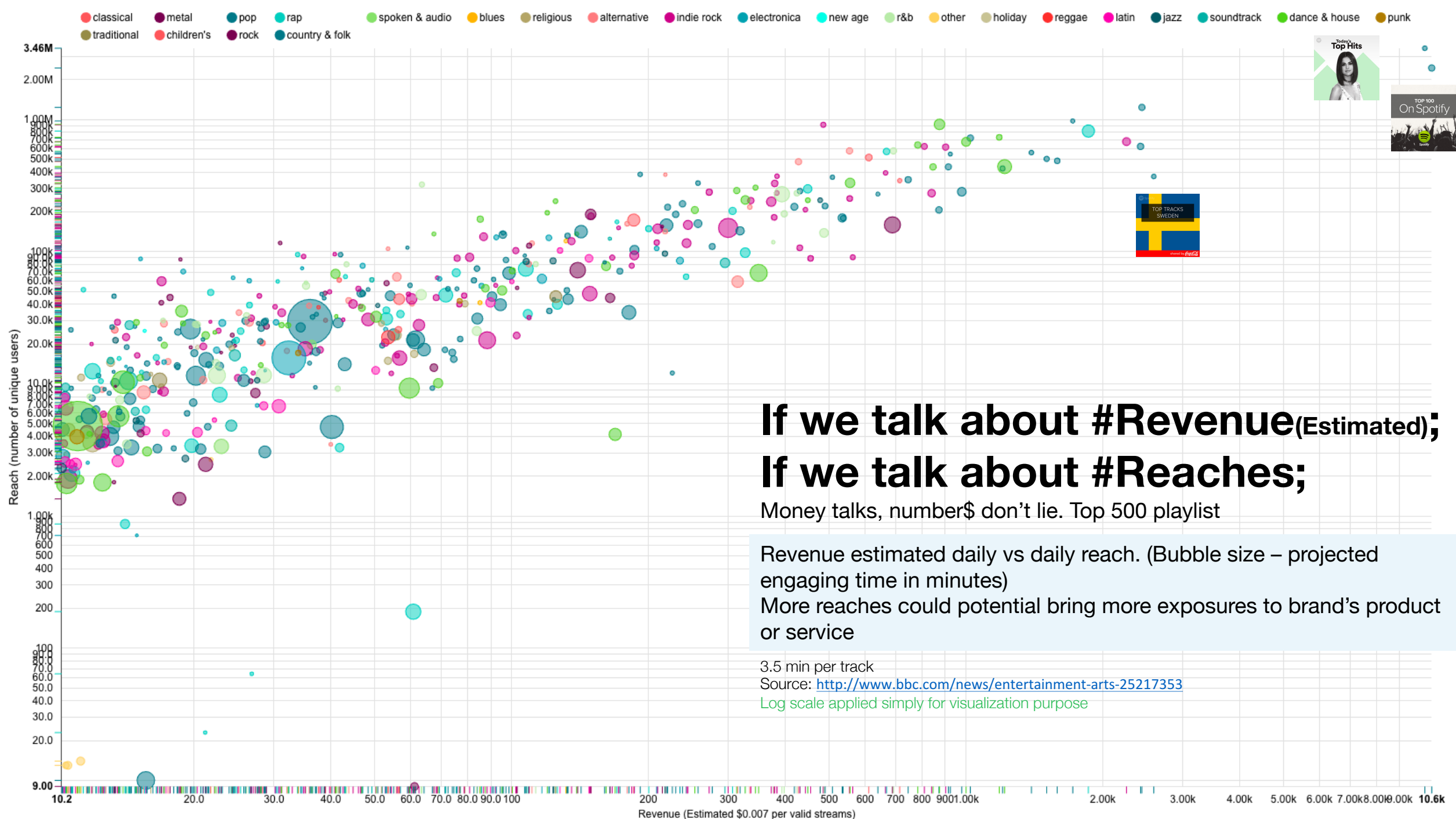**All scripts, code, and data is available in the following repo**
https://github.com/thehongwu/playlist_summary_analysis
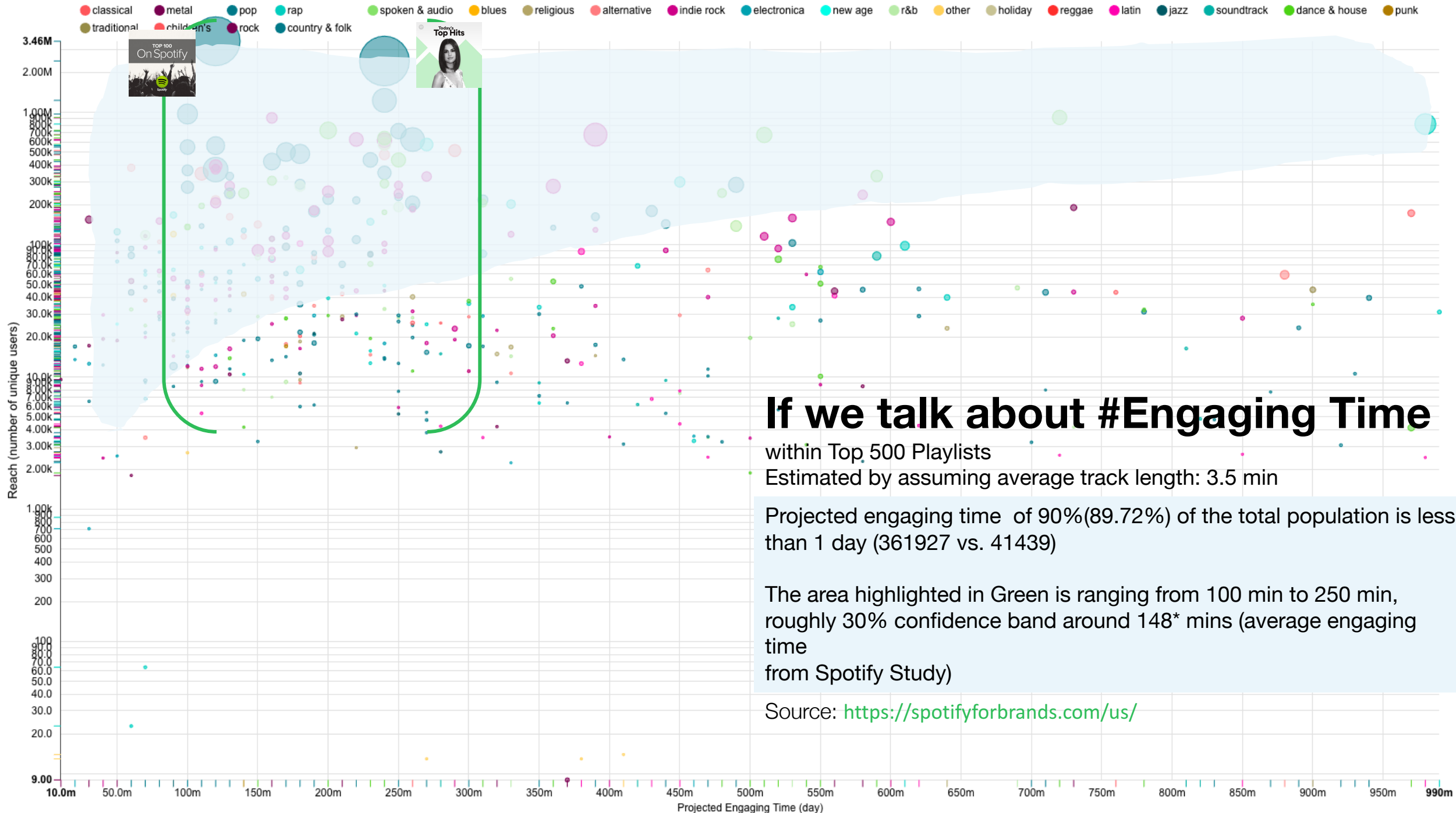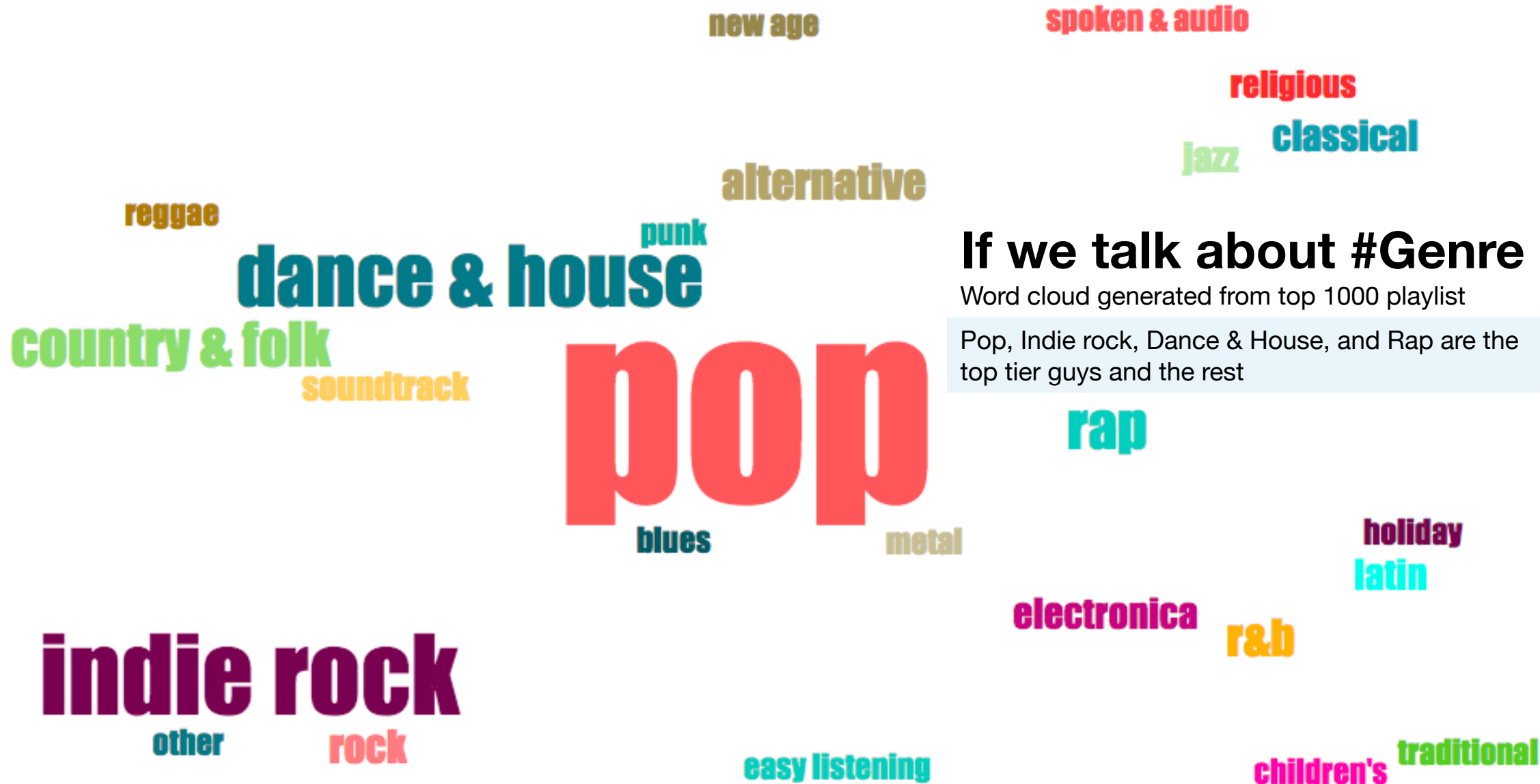Will delete after interview

Technical Sheet – Audience: Data/Tech Team

A little bit on Data Processing

- mysql
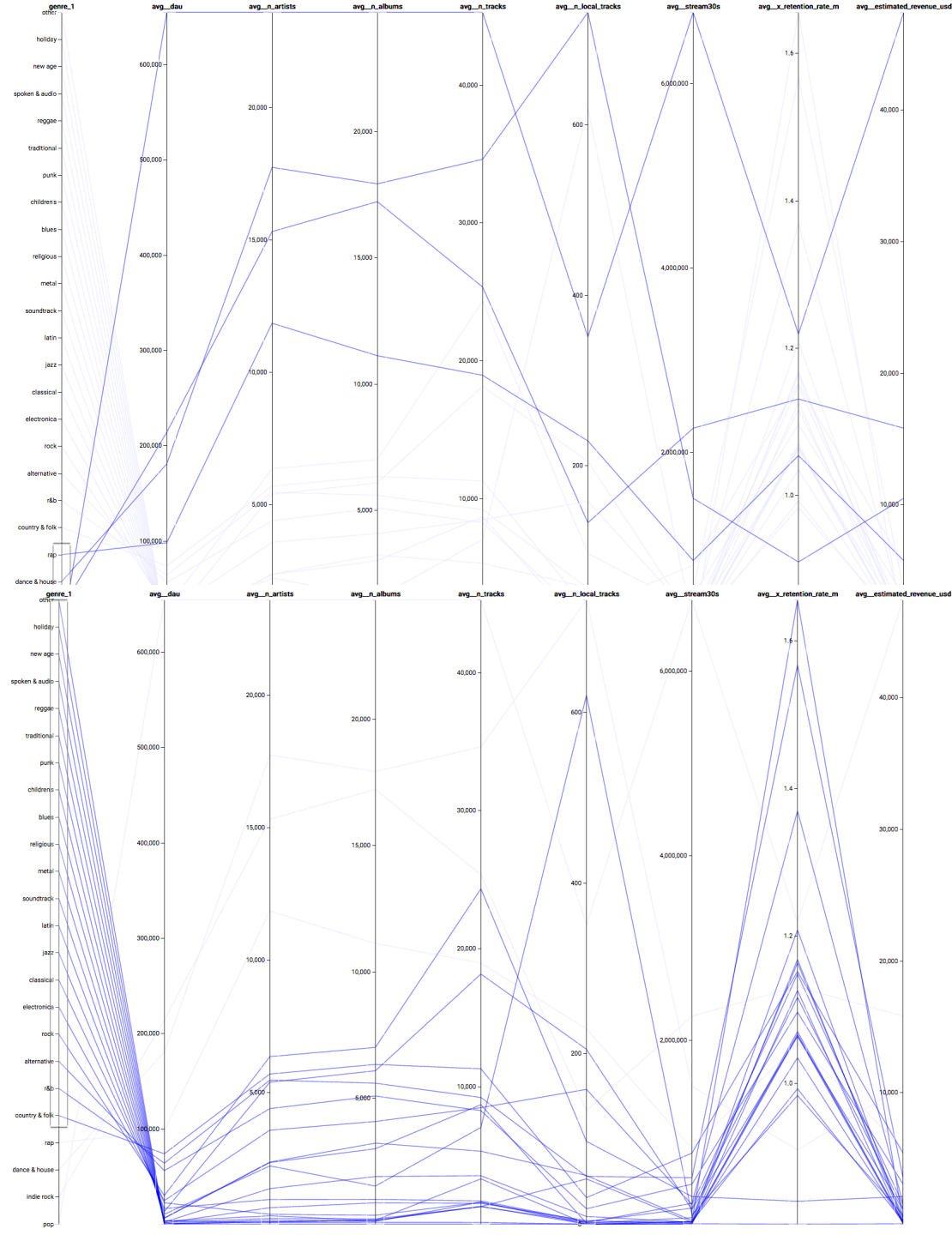- Spotify API: https://developer.spotify.com/web-api/console/get-playlist/#complete
- Python, superset
- NLTK

Legend (top):
classical, metal, pop, rap, spoken & audio, blues, religious, alternative, indie rock, electronica, new age, r&b, other, holiday, reggae, latin, jazz, soundtrack, dance & house, punk, traditional, children's, rock, country & folk

Y-axis: Reach (number of unique users)
Values: 3.46M, 2.00M, 1.00M, 900k, 800k, 700k, 600k, 500k, 400k, 300k, 200k, 100k, 90.k, 80.k, 70.k, 60.k, 50.k, 40.k, 30.k, 20.k, 10.0k, 9.00k, 8.00k, 7.00k, 6.00k, 5.00k, 4.00k, 3.00k, 2.00k, 1.00k, 900, 800, 700, 600, 500, 300, 200, 100, 90.0, 80.0, 70.0, 60.0, 50.0, 40.0, 30.0, 20.0, 9.00

X-axis: Revenue (Estimated $0.007 per valid streams)
Values: 10.2, 20.0, 30.0, 40.0, 50.0, 60.0, 70.0, 80.0, 90.0, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1.00k, 2.00k, 3.00k, 4.00k, 5.00k, 6.00k, 7.00k, 8.00k, 9.00k, 10.6k

If we talk about #Revenue(Estimated);
If we talk about #Reaches;
Money talks, number$ don't lie. Top 500 playlist

Revenue estimated daily vs daily reach. (Bubble size – projected engaging time in minutes)
More reaches could potential bring more exposures to brand's product or service

3.5 min per track
Source: http://www.bbc.com/news/entertainment-arts-25217353
Log scale applied simply for visualization purpose

Legend (top): classical · metal · pop · rap · spoken & audio · blues · religious · alternative · indie rock · electronica · new age · r&b · other · holiday · reggae · latin · jazz · soundtrack · dance & house · punk · traditional · children's · rock · country & folk

Y-axis: Reach (number of unique users)

X-axis: Projected Engaging Time (day)

If we talk about #Engaging Time

within Top 500 Playlists
Estimated by assuming average track length: 3.5 min

Projected engaging time of 90%(89.72%) of the total population is less than 1 day (361927 vs. 41439)

The area highlighted in Green is ranging from 100 min to 250 min, roughly 30% confidence band around 148* mins (average engaging time
from Spotify Study)

Source: https://spotifyforbrands.com/us/

new age

spoken & audio

religious

classical

jazz

reggae

alternative

punk

dance & house

If we talk about #Genre

Word cloud generated from top 1000 playlist

Pop, Indie rock, Dance & House, and Rap are the top tier guys and the rest

country & folk

soundtrack

pop

rap

blues

metal

holiday

latin

electronica

r&b

indie rock

other

rock

easy listening

children's

traditional

If we talk about #Genre

Parallel coordinates generated from top 1000 playlist
Which verifies the result from word cloud:

"More Active user, more valid stream, more engaging time"

DAU is a "good splitor" which indictors "Potential Two Clusters" in terms of common "metrics":
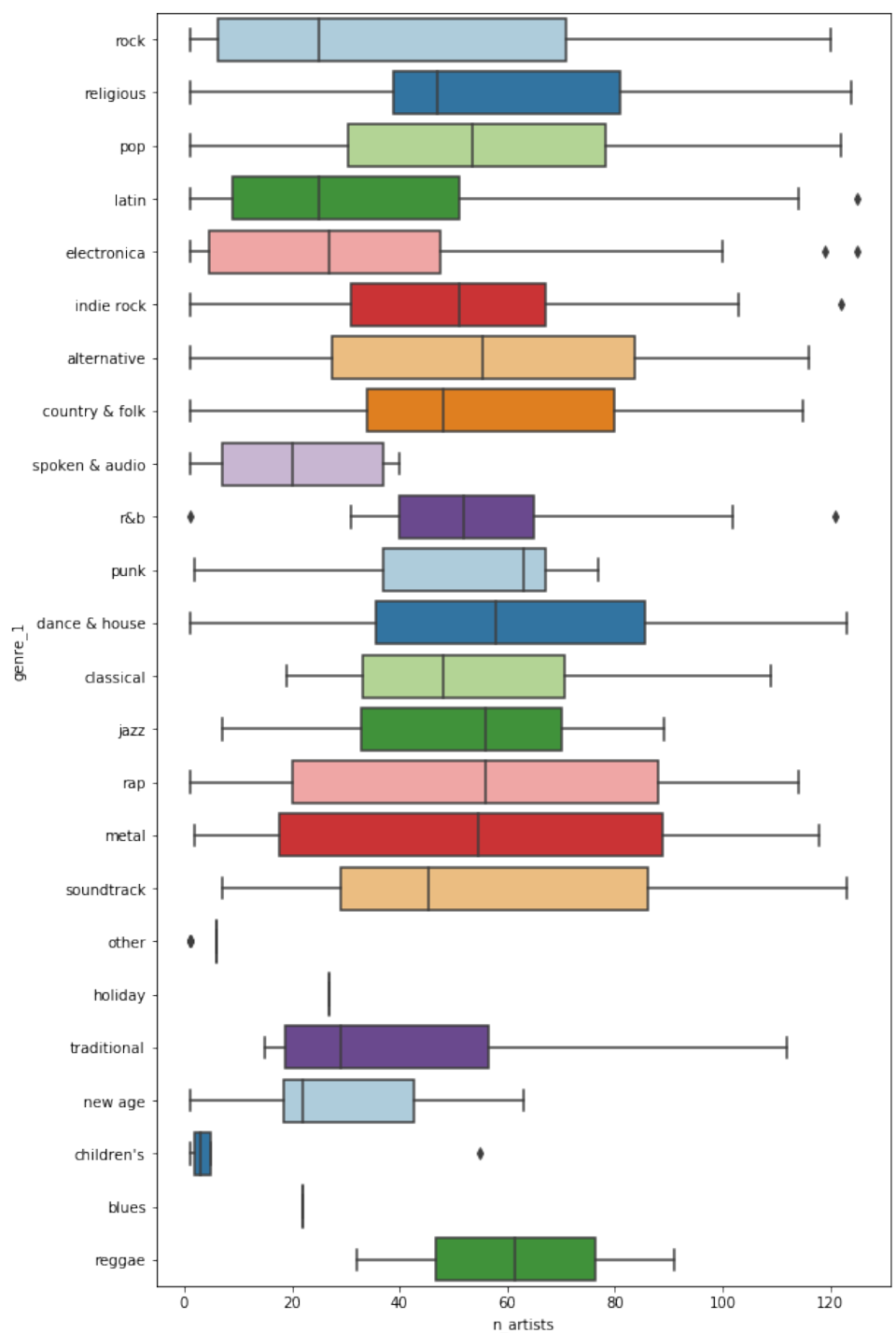
- {Pop, Indie Rock, Dance & House, Rap}
- Vs. {The rest}

tender

peaceful   sensual

lively

sentimental   defiant

empowering   upbeat   excited   yearning

somber   romantic

easygoing   fiery

If we talk about #Mood
Word cloud generated from top 1000 playlist

urgent   energizing

other

Top moods:
- Excited
- defiant
- yearning
- melancholy

melancholy

gritty   sophisticated

serious

aggressive   rowdy   cool
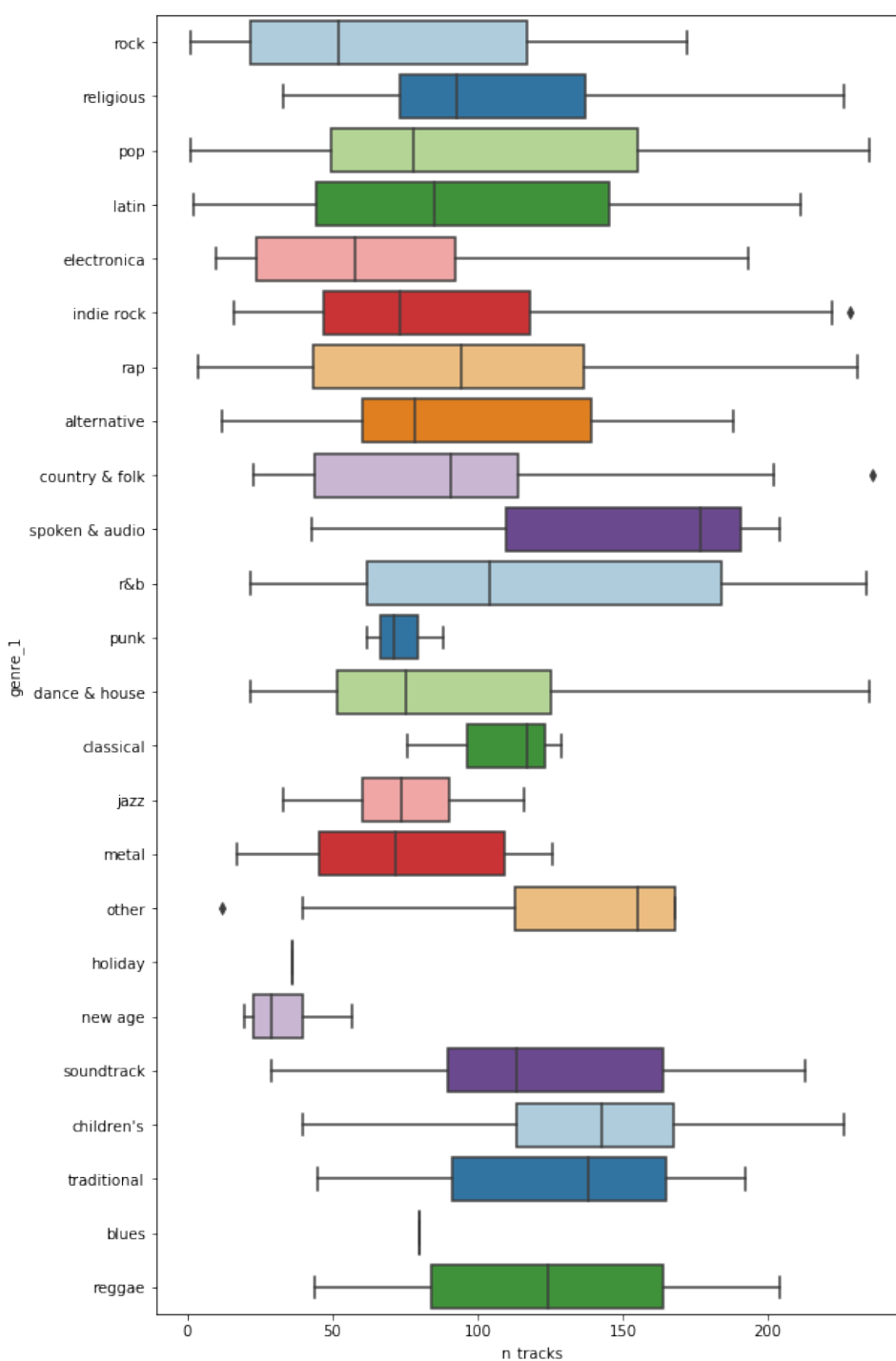
If we talk about #Tracks/Artists

Boxplot generate from top 1000 playlist
with extreme/anomalies data taken out

We can generalize what a playlist looks like
from this perspective, some highlights:
pop: 81 tracks - 60 artist similar number of album
indie rock: 78 tracks - 58 artist similar number of album
dance & house: 83 tracks - 65 artist similar number of album
rap: 95 tracks - 75 artist similar number of album

A formula of a "successful" playlist leading to high engaging and revenue-generation.

NOTE:
Purely from this datasets, and data/metrics perspective along with assumptions

A playlist of **Indie Rock** **Pop** **Dance & House** **Rap** around **80ish** tracks by **65ish** artists bringing **Excited** **Defiant** **Yearning** **Melancholy**

One more thing...

If we talk about #MUSIC
If we talk about #USERS
If we talk about #INTEREST

If we talk about #Tastes

With tokens available, we are able to extract top 5 taste from top 1000 playlists

- {countries, rock, best, parties, mix}
- {pop, hit, music, summer, folk}
- {music, Christian, new, parties, update}
- {playlist, ultimate, workout, best, summer}
- {song, love, classic, new, countries}
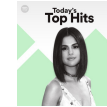
Majority of the active users could be characterized as "Explorers": loves country music, party and its summer!!! which possibly opens new opportunities for Brands and advertiser to do targeted marketing, trying new stuffs, new restaraunts, things related to party, travel, concert etc.

"What we "REALLY" talk about
when we talk about successful playlist"

Affinity Between Content and Audience