Business Programming (using Python)

Zhaohu (Jonathan) Fan

Information Technology Management

Scheller College of Business

Georgia Institute of Technology

November 21, 2023

What this presentation is about

- Understanding the Users' behavior on **Spotify**.

Our research question is

- How have the interests of **Spotify** users changed over time?

Spotify's songs data

- **Spotify** original data includes 32,833 tracks and 23 attributes.

Spotify's songs data (cont'd)

- **Spotify** cleaned data includes 28,352 tracks and 23 attributes.

Playlist Genre

- **6** Main categories
  - **Latin**
  - **Pop**
  - **R&B(Rhythm & Blues)**
  - **Rap**
  - **Rock**
  - **EDM (Electronic Dance Music)**

Characteristics of Interests

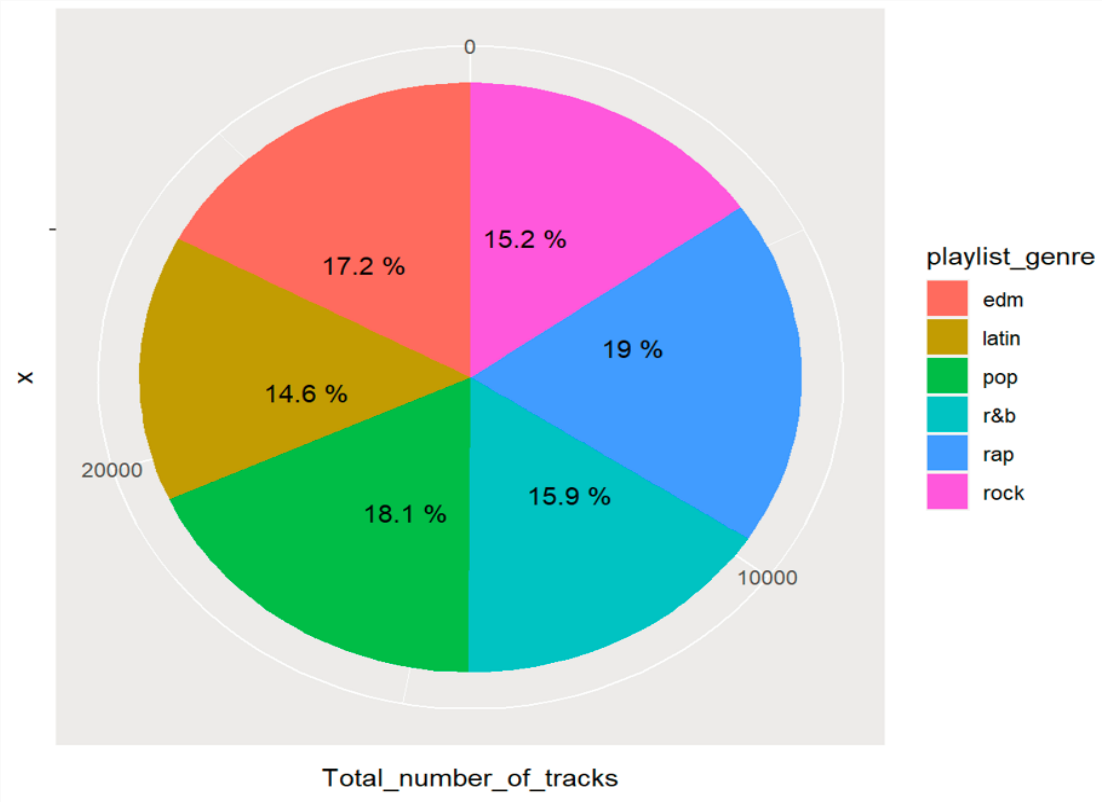
- **Loudness**: quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typical range between -60 and 0 db.
- **Speechiness**: detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value.
- **Acousticness**: measures from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.

Characteristics of Interests (cont'd)

- **Track popularity**: song popularity (0-100) where higher is better.
- **Danceability** : 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.
- **Energy**: represents a perceptual measure of intensity and activity. Energy is a measure from 0.0 to 1.0. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale.

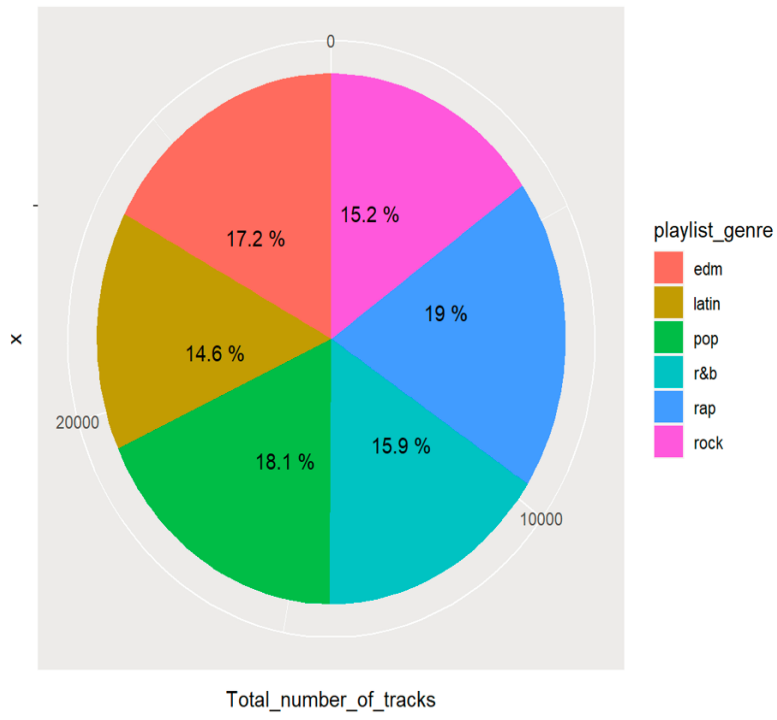
Exploratory Data Analysis

- The release of tracks by genres



Exploratory Data Analysis

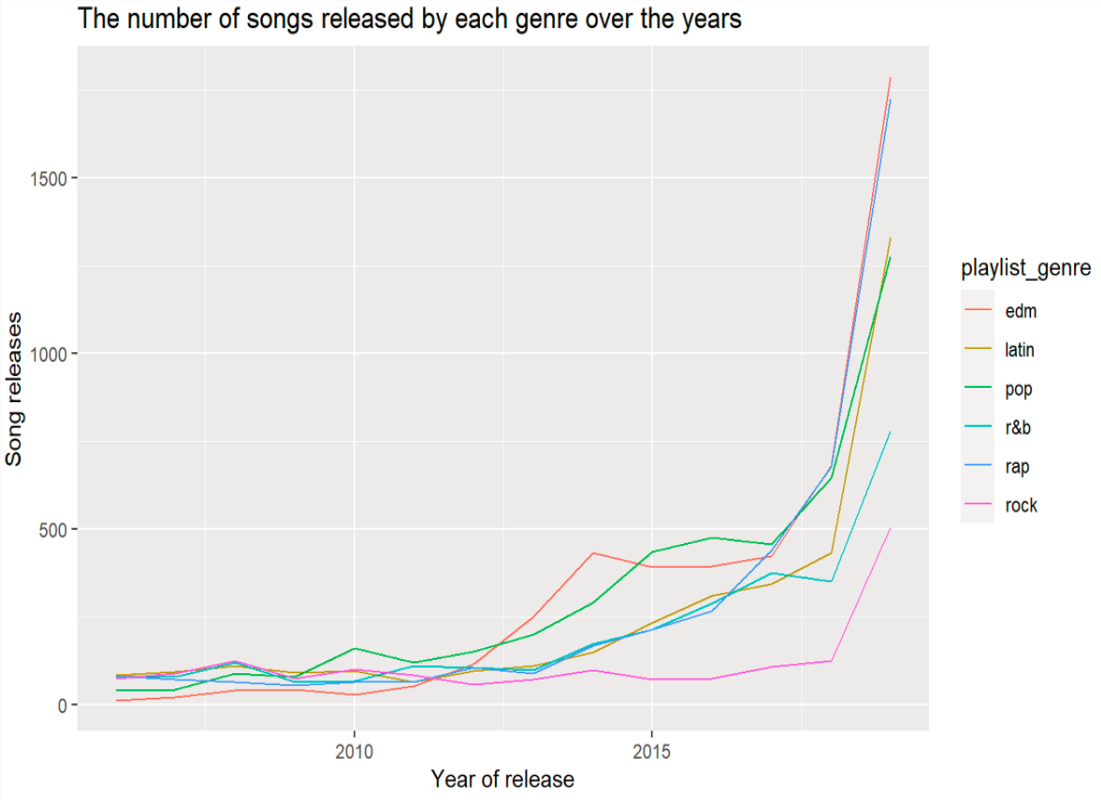
- The release of tracks by genres



- **Observations**
  - Across all playlist genres, **rap music** has the greatest number of tracks.
  - **Latin music** has the lowest number of tracks of all playlist genres.
  - **There are no genres that are particularly prominent in the majority**.

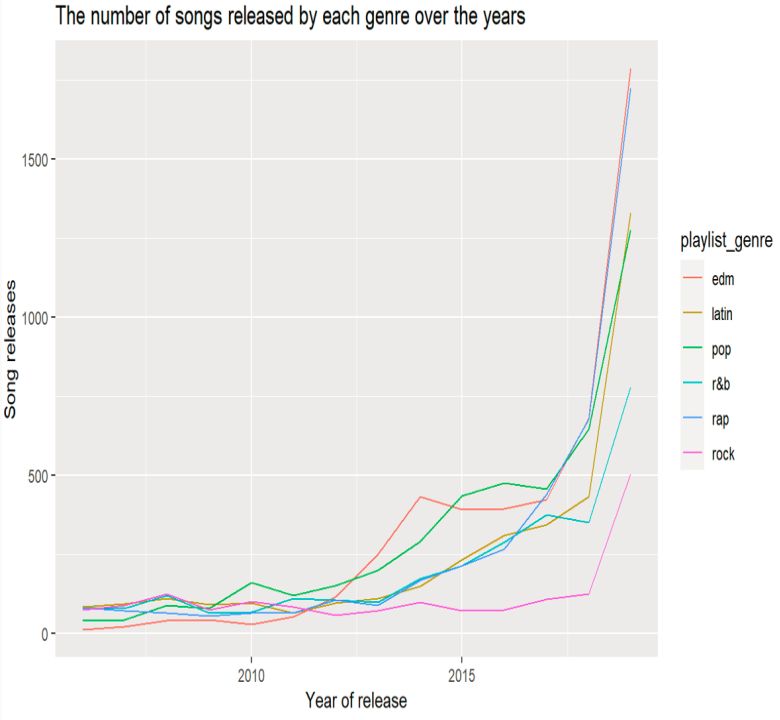
Exploratory Data Analysis (cont'd)

- Number of tracks released between 2005 and 2019 based on genres



Exploratory Data Analysis (cont'd)

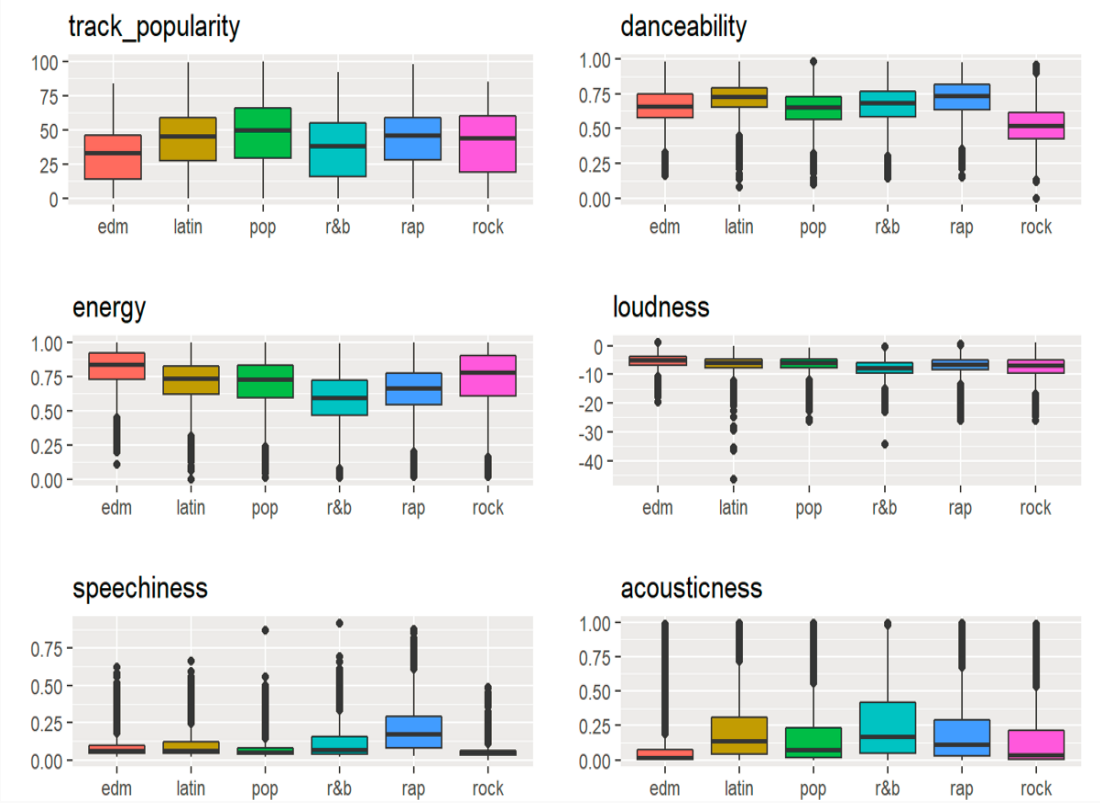
- Number of tracks released between 2005 and 2019 based on genres



- **Observations**
  - Among all other genres, **rock music** releases the least number of songs each year.
  - Until 2010, **EDM** wasn't that popular, however the number of EDM songs released increased dramatically after 2013 and reached its peak in 2019.

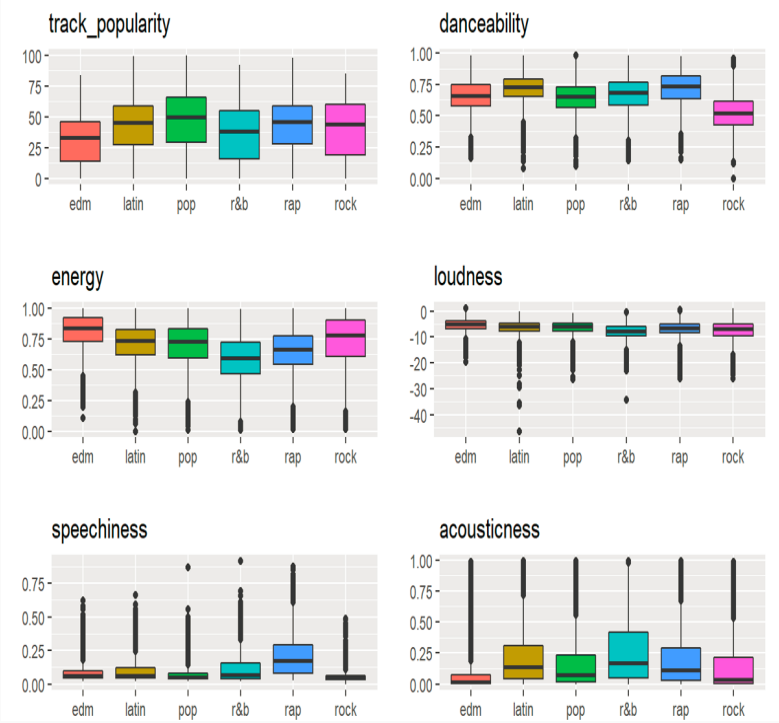
Exploratory Data Analysis (cont'd)

- Key characteristics of interests by genre.



Exploratory Data Analysis (cont'd)

- Key characteristics of interests by genre.



- **Observations**
  - **Pop music** enjoys the highest popularity among all genres.
  - **Rap music** accounts for the highest degree of danceability.
  - Among all genres of music, **Electronic dance music (EDM)** has the highest energy and loudness.

Insights

- We have briefly examined the behavior patterns of users.

Insights (cont'd)

- Spotify's library of tracks consists of tracks from a wide variety of genres, but the most popular tracks are from the pop genre.)
- Rap music is the most danceable genre.

Implications

- We propose a preliminary recommendation system by using cluster analysis.

Implications (cont'd)

- Our users can thus be suggested songs based on the songs they have previously listened to.

Implications (cont'd)

- We expect that this will lead to a higher rate of customer retention as users will be able to access similar playlists.

Next step(s)

- We plan to improve the accuracy and efficiency of our proposed recommendation system.

Next step(s) (cont'd)

- Particularly, we will incorporate domain experts' knowledge and the core ideas of discrete choice modelling into our recommendation system, since the types of genres are nominal variables.

Questions?