DATA ANALYTICS CAPSTONE I

Jenny Duong

Dataset

Spotify Tracks Genres

[SOURCE]

This dataset provides comprehensive information about Spotify tracks encompassing a diverse collection of 125 genres. It has been compiled and cleaned using Spotify's Web API and Python. Presented in CSV format, this dataset is easily accessible and amenable to analysis. The dataset comprises multiple columns, each representing distinctive audio features associated with individual tracks.

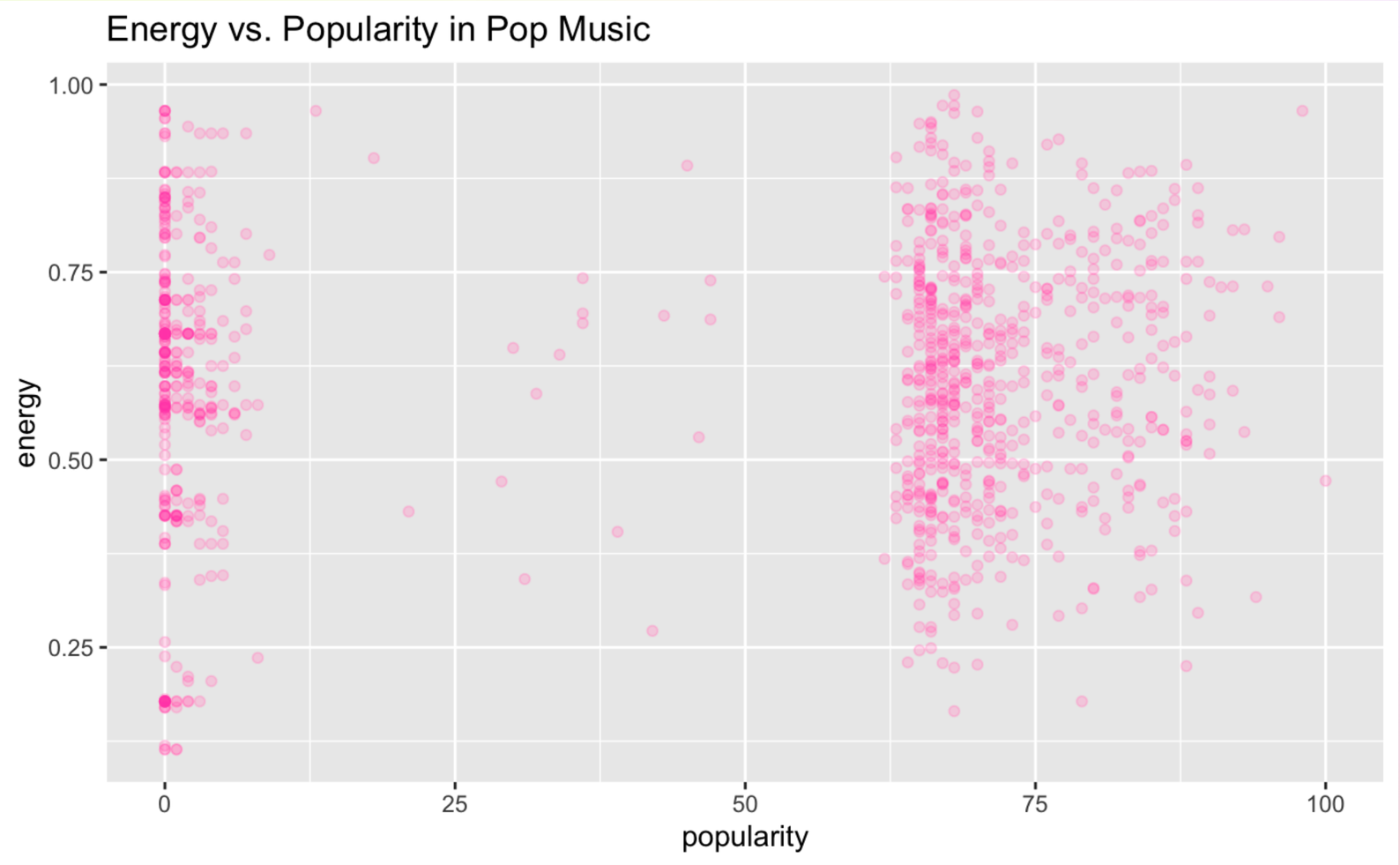
Data Columns

ARTISTS - THE NAME(S) OF THE ARTIST(S) ASSOCIATED WITH THE TRACK. (STRING)
ALBUM_NAME - THE NAME OF THE ALBUM THAT THE TRACK BELONGS TO. (STRING)
TRACK_NAME - THE NAME OF THE TRACK. (STRING)
POPULARITY - THE POPULARITY SCORE OF THE TRACK ON SPOTIFY, RANGING FROM 0 TO 100. (INTEGER)
DURATION_MS - THE DURATION OF THE TRACK IN MILLISECONDS. (INTEGER)
EXPLICIT - A BOOLEAN VALUE INDICATING WHETHER THE TRACK CONTAINS EXPLICIT CONTENT. (BOOLEAN)
DANCEABILITY - A SCORE RANGING FROM 0-1 THAT REPRESENTS HOW SUITABLE A TRACK IS FOR DANCING BASED ON VARIOUS MUSICAL ELEMENTS. (FLOAT)
ENERGY - A MEASURE OF THE INTENSITY AND ACTIVITY OF A TRACK, RANGING FROM 0 TO 1. (FLOAT)
KEY - THE KEY OF THE TRACK REPRESENTED BY AN INTEGER VALUE. (INTEGER)
LOUDNESS - THE LOUDNESS OF THE TRACK IN DECIBELS (DB). (FLOAT)
MODE - THE TONAL MODE OF THE TRACK, REPRESENTED BY AN INTEGER VALUE (0 FOR MINOR, 1 FOR MAJOR). (INTEGER)
SPEECHINESS - A SCORE RANGING FROM 0 TO 1 THAT REPRESENTS THE PRESENCE OF SPOKEN WORDS IN A TRACK. (FLOAT)
ACOUSTICNESS - A SCORE RANGING FROM 0 TO 1 THAT REPRESENTS THE EXTENT TO WHICH A TRACK POSSESSES AN ACOUSTIC QUALITY. (FLOAT)
INSTRUMENTALNESS - A SCORE RANGING FROM 0 TO 1 THAT REPRESENTS THE LIKELIHOOD OF A TRACK BEING INSTRUMENTAL. (FLOAT)
LIVENESS - A SCORE RANGING FROM 0-1 THAT REPRESENTS THE PRESENCE OF AN AUDIENCE DURING THE RECORDING OR PERFORMANCE OF A TRACK. (FLOAT)
VALENCE - A SCORE RANGING FROM 0 TO 1 THAT REPRESENTS THE MUSICAL POSITIVENESS CONVEYED BY A TRACK. (FLOAT)
TEMPO - THE TEMPO OF THE TRACK IN BEATS PER MINUTE (BPM). (FLOAT)
TIME_SIGNATURE - THE NUMBER OF BEATS WITHIN EACH BAR OF THE TRACK. (INTEGER)
TRACK_GENRE - THE GENRE OF THE TRACK. (STRING)

Hypothesis #1

ENERGY AND POPULARITY : TRACKS WITH HIGH POPULARITY WILL ALSO HAVE A HIGH ENERGY RATING.

Exploratory Analysis

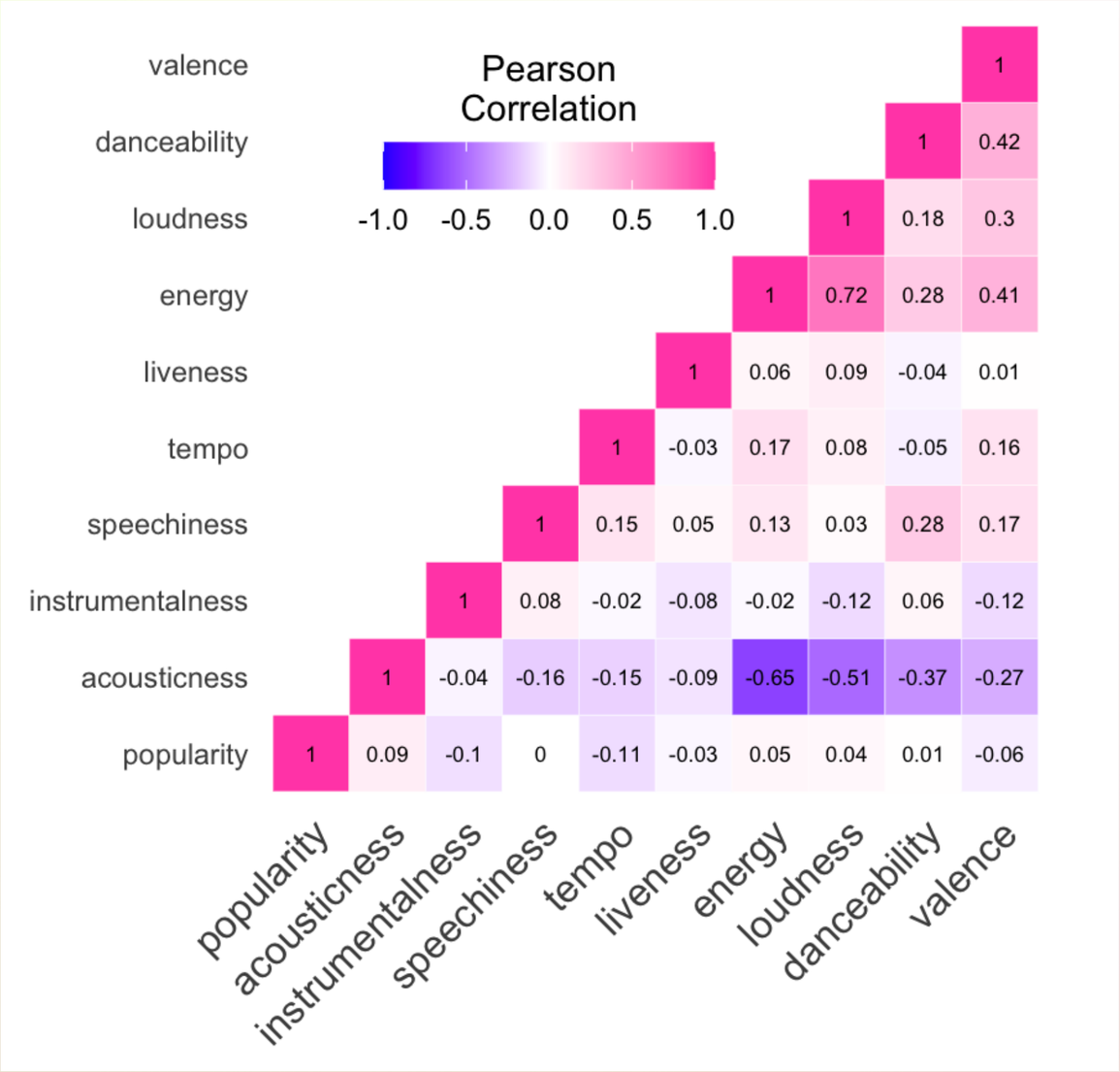


I started my analysis with a scatter plot to see if there was any correlation between a track's energy rating and its popularity.

I filtered out only tracks in the "pop" genre since it was the most prolific genre.

There didn't seem to be any correlation.

Exploratory Analysis



From there, I got curious to see if there was any correlation between ANY of the categories.

Again, filtering out only “pop” tracks for consistency, I created a Pearson Correlation chart.

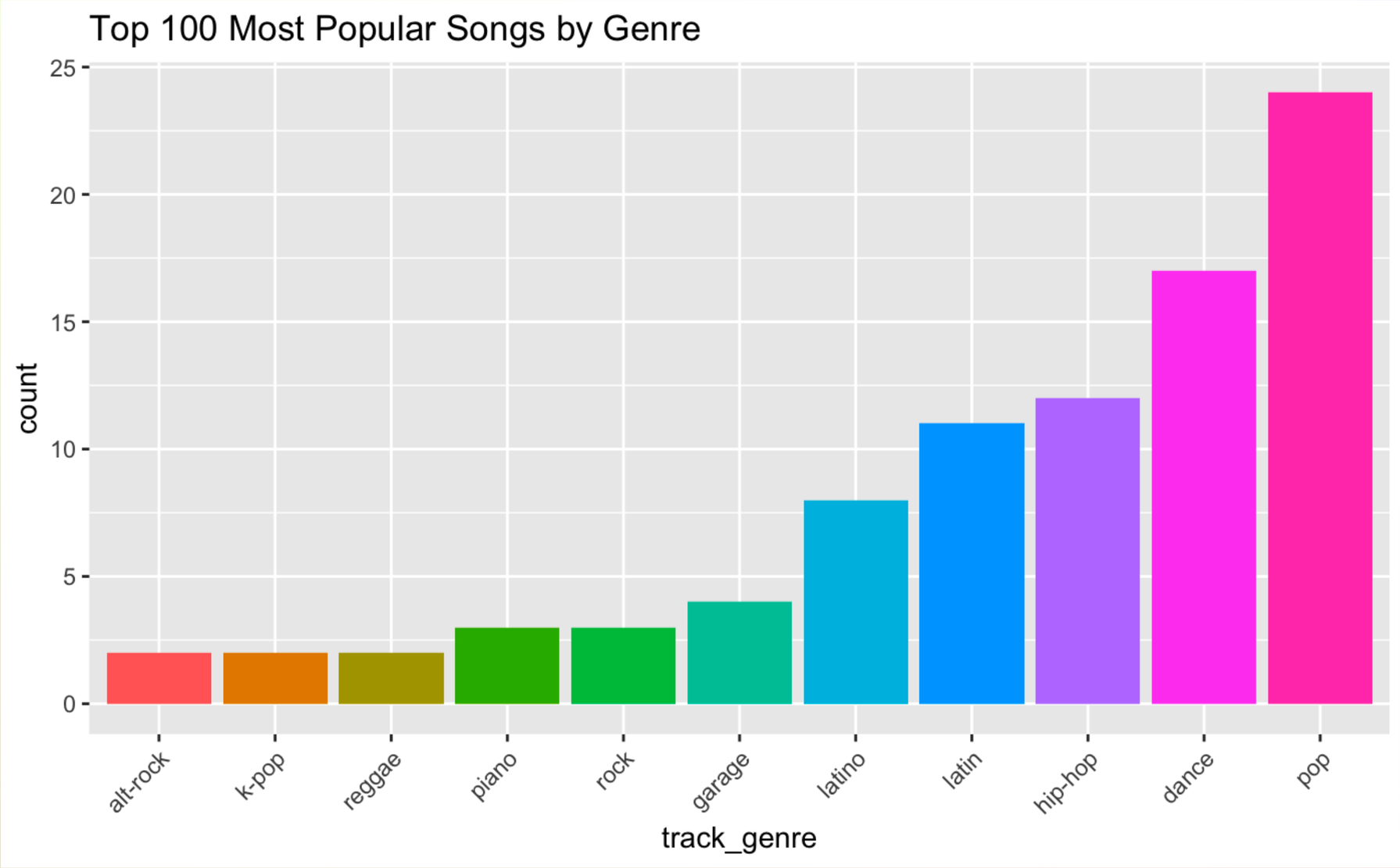
The highest correlation was between energy and loudness. Acousticness and energy had the highest negative correlation. There was also a strong correlation between the danceability of a song and its valence (how positive it sounded).

“Pop” music is sometimes synonymous with popular music, so I decided to look into the attributes of popular songs.

Hypothesis #2

THE MOST POPULAR GENRE WILL BE POP, AND THE MOST POPULAR POP TRACKS WILL HAVE HIGH VALUES OF DANCEABILITY.

Exploratory Analysis

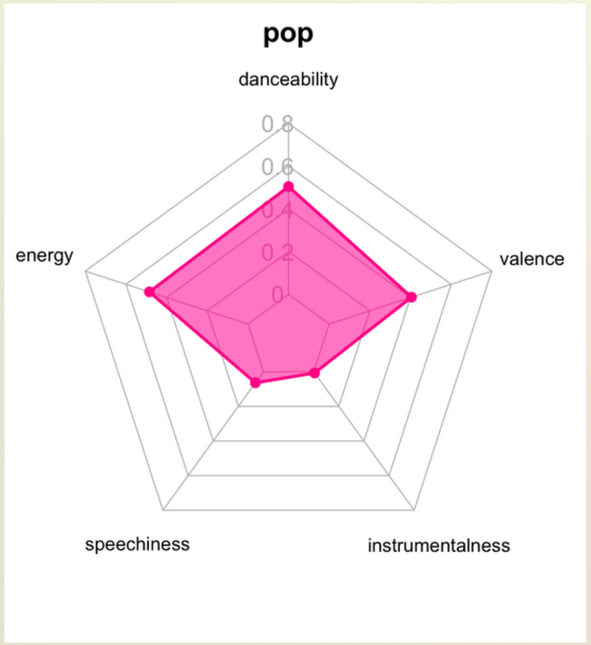


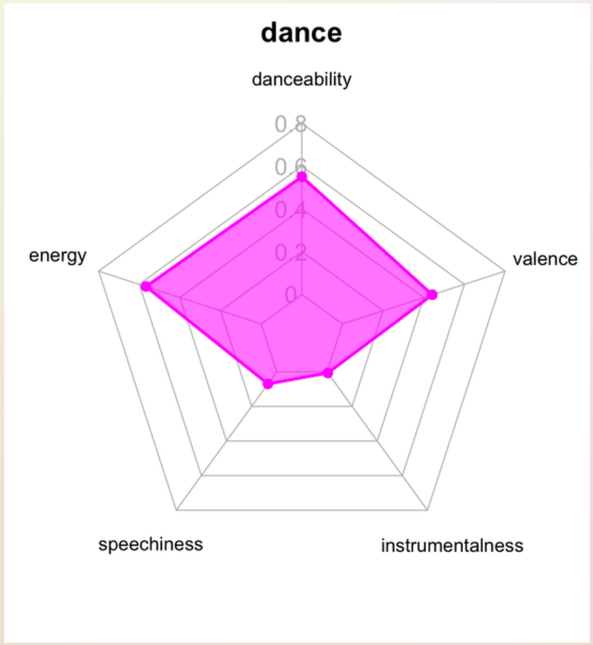
I grabbed the top 100 most popular songs, then organized them by track genre. Pop was the most popular, followed by dance, hip-hop, then latin.

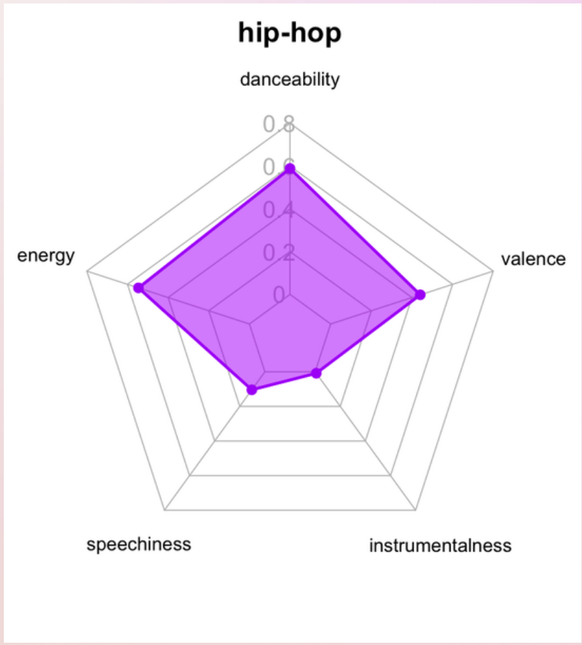
Exploratory Analysis

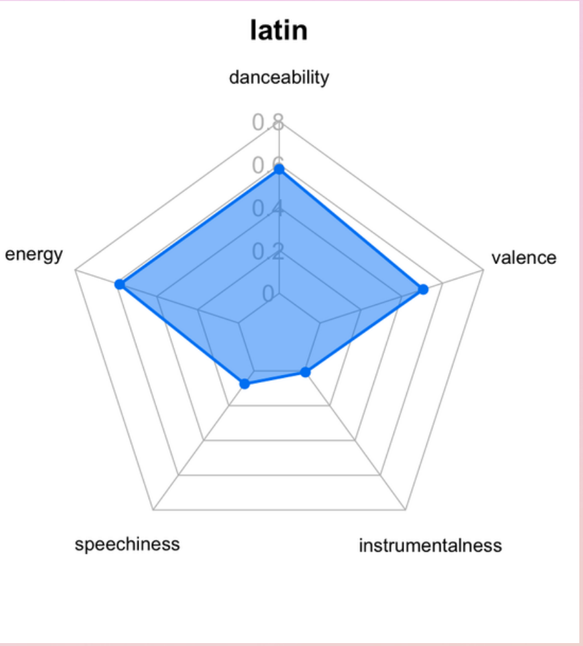
I then used a radar chart to look at the average qualities of the four most popular genres. They were all very similar...

According to the Wikipedia, “identifying factors of pop music usually include repeated choruses and hooks [...] and rhythms or tempos that can be easily danced to. Much pop music also borrows elements from other styles such as rock, urban, dance, Latin, and country.”

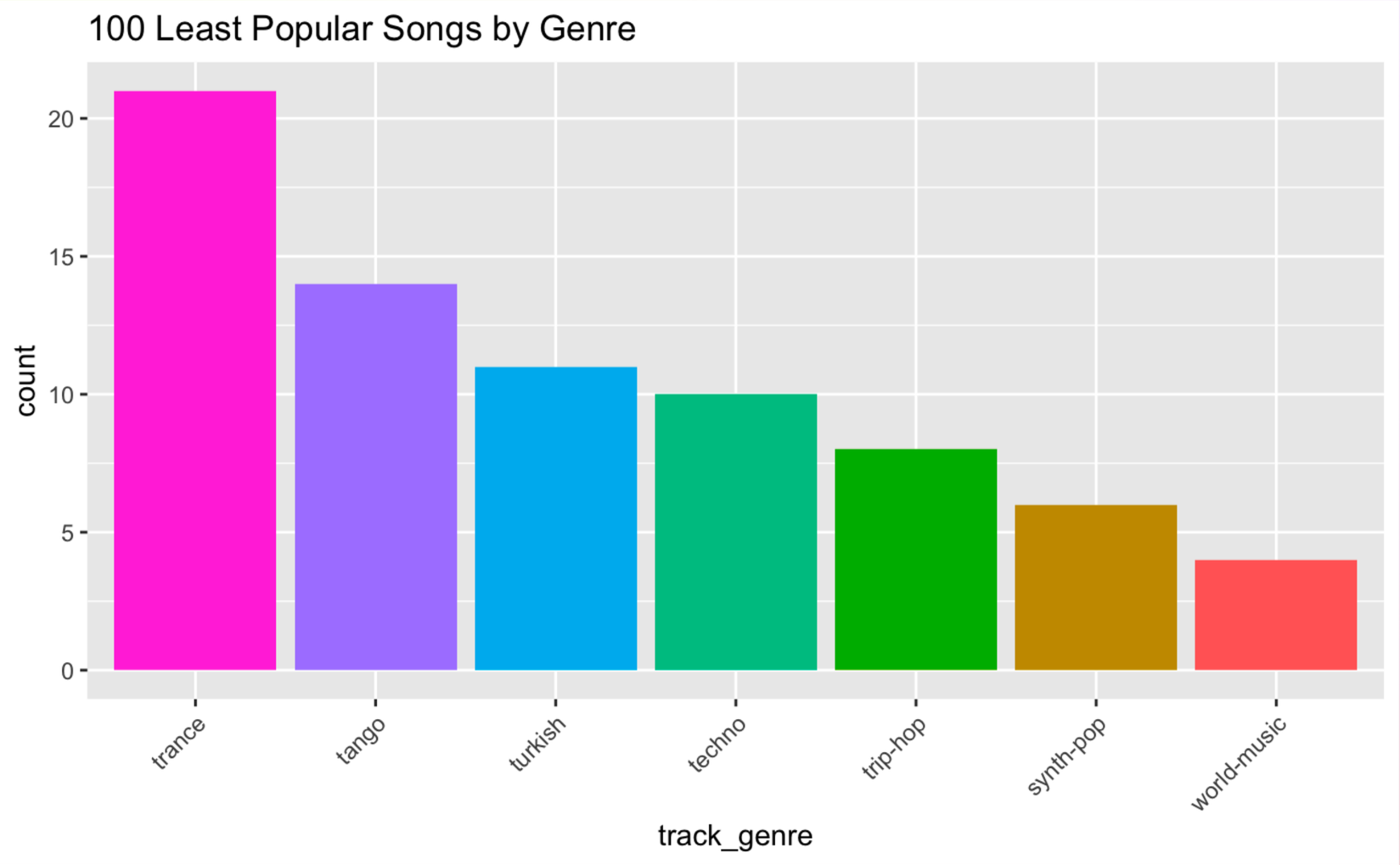








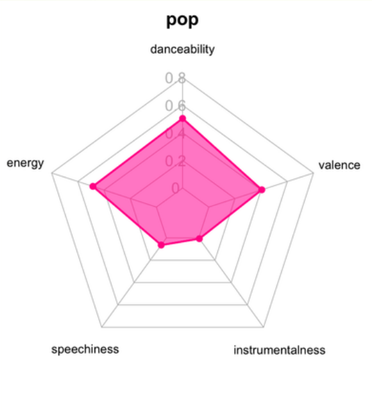
Exploratory Analysis

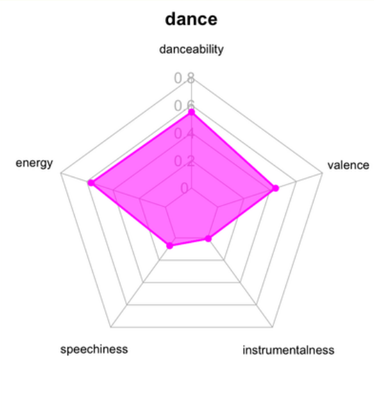


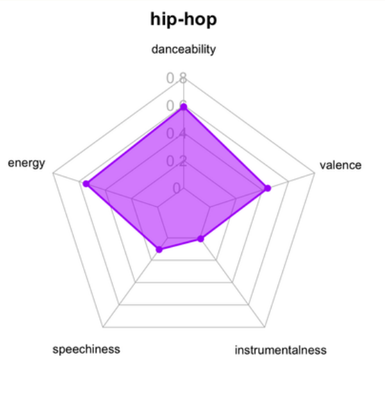
I decided to look at the least popular tracks to see if they had anything in common.

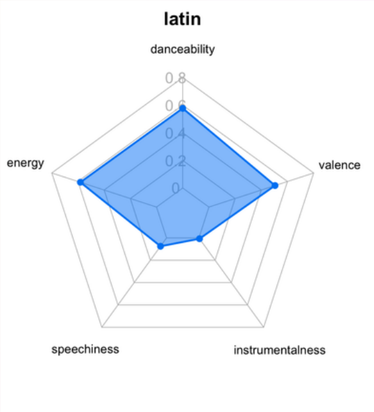
Trance, tango, turkish, and techno rounded out the bottom four genres.

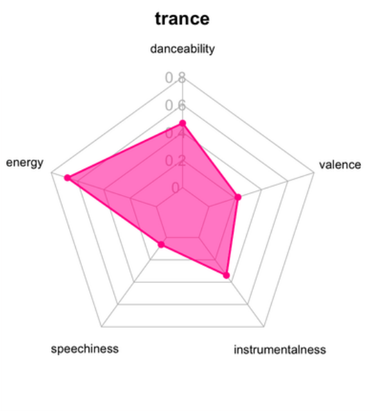
Exploratory Analysis

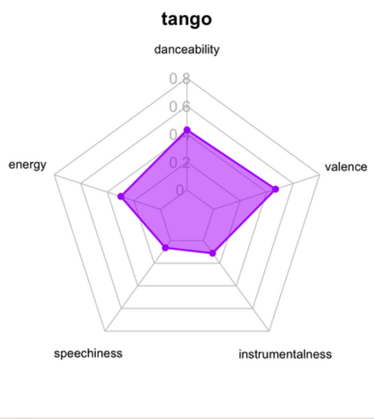


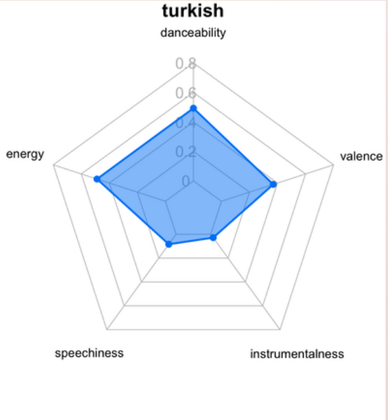


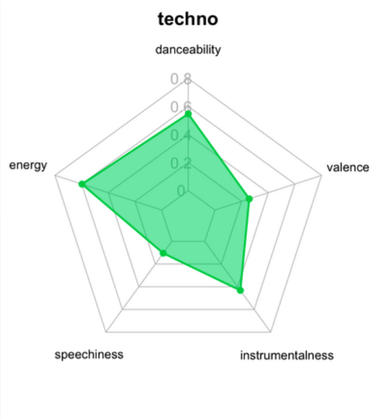












I then used radar charts to look at the average qualities of the four least popular genres.

Besides turkish music, they all were different shapes from the popular genres. Instrumentalness was notably more prevalent.

It was very interesting to look at the various elements of the most and least popular tracks. It was apparent that all the most popular genres shared very similar traits, while the least popular genres were more eclectic.

People seem to love a high energy, positive and danceable bop.

Conclusion

End