Understanding Emotional Themes in Spotify Playlists: Final Report

By Lawrence Ewald

Objective

Music is an important part of everyone’s lives and a poet might even say we were gifted to ears to hear and arrange melodies to express our own thoughts and feelings. In the 21st century, streaming platforms like Spotify have risen to be a major way people go about listening to music. My project aims to examine the frequency of songs within playlists and identifying if there is a positive or negative theme across the most popular spotify playlists, asking What is the dominant emotional theme across a majority of spotify playlists?

Datasets

- Spotify Million Song Dataset [SMSD] (last updated in 2023)

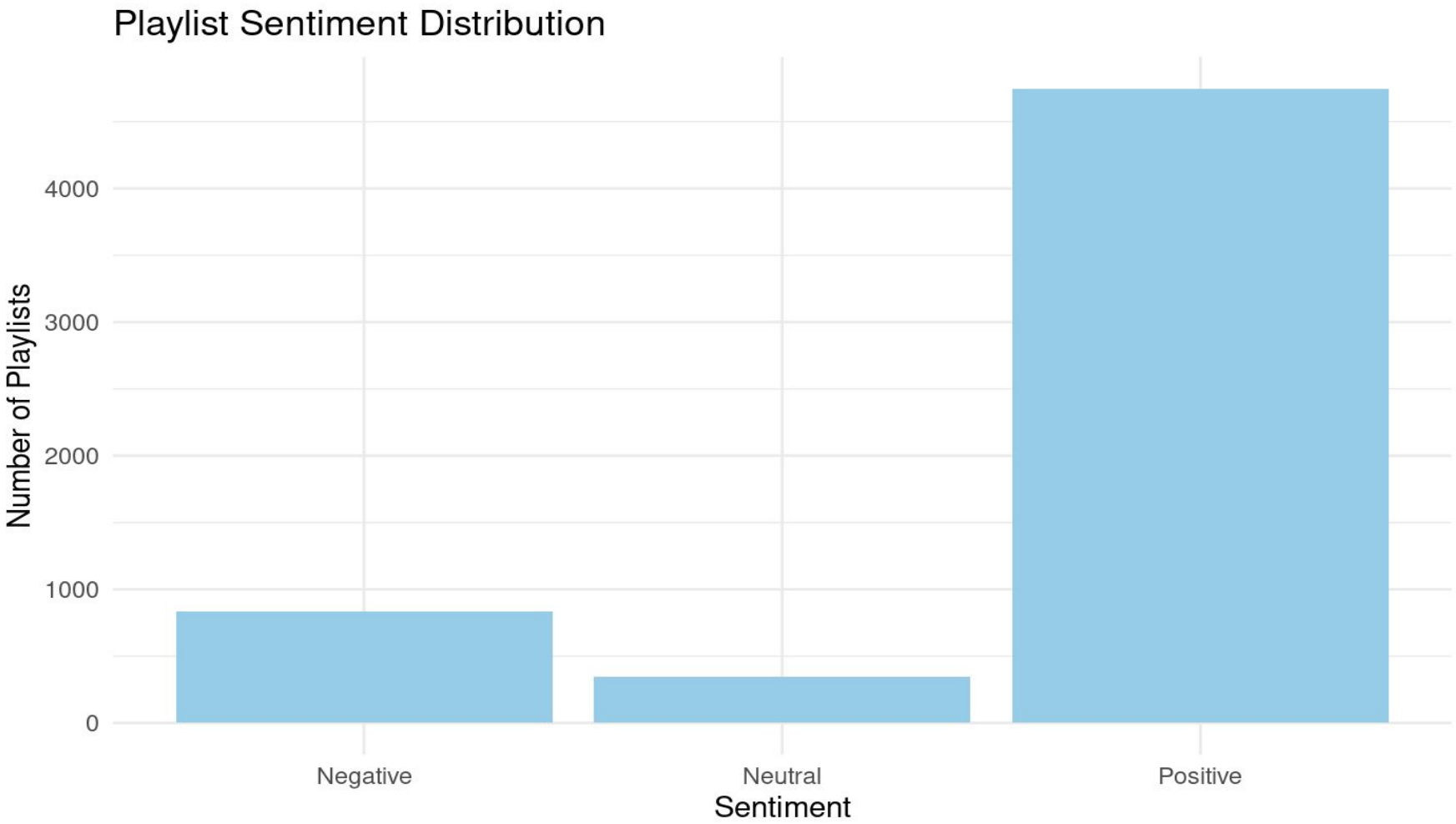
This dataset contains a million songs which will provide a basis of seeing how frequently these songs appear within playlists.

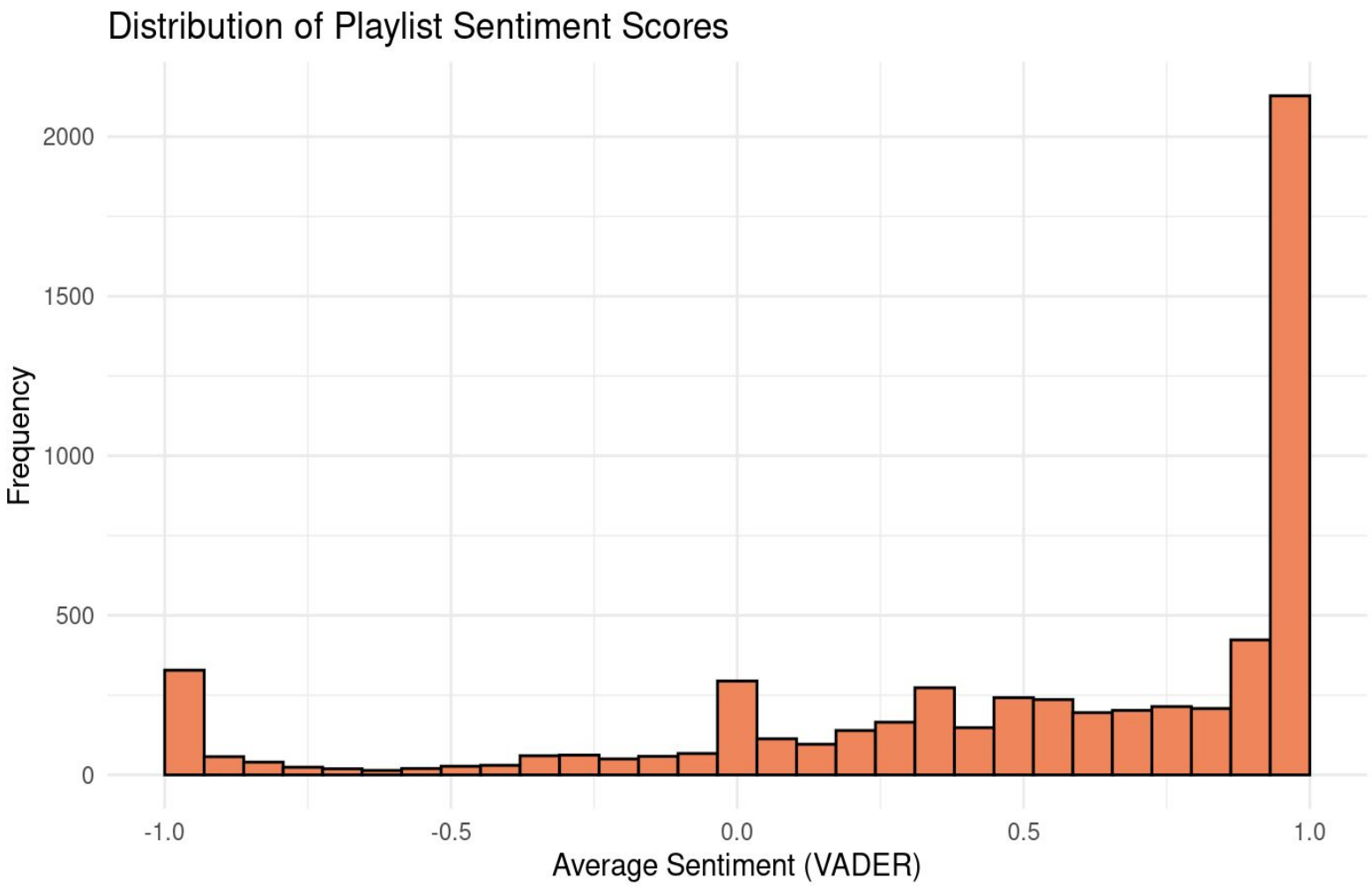
- Spotify Million Playlist Database [SMPD] (created in 2020)

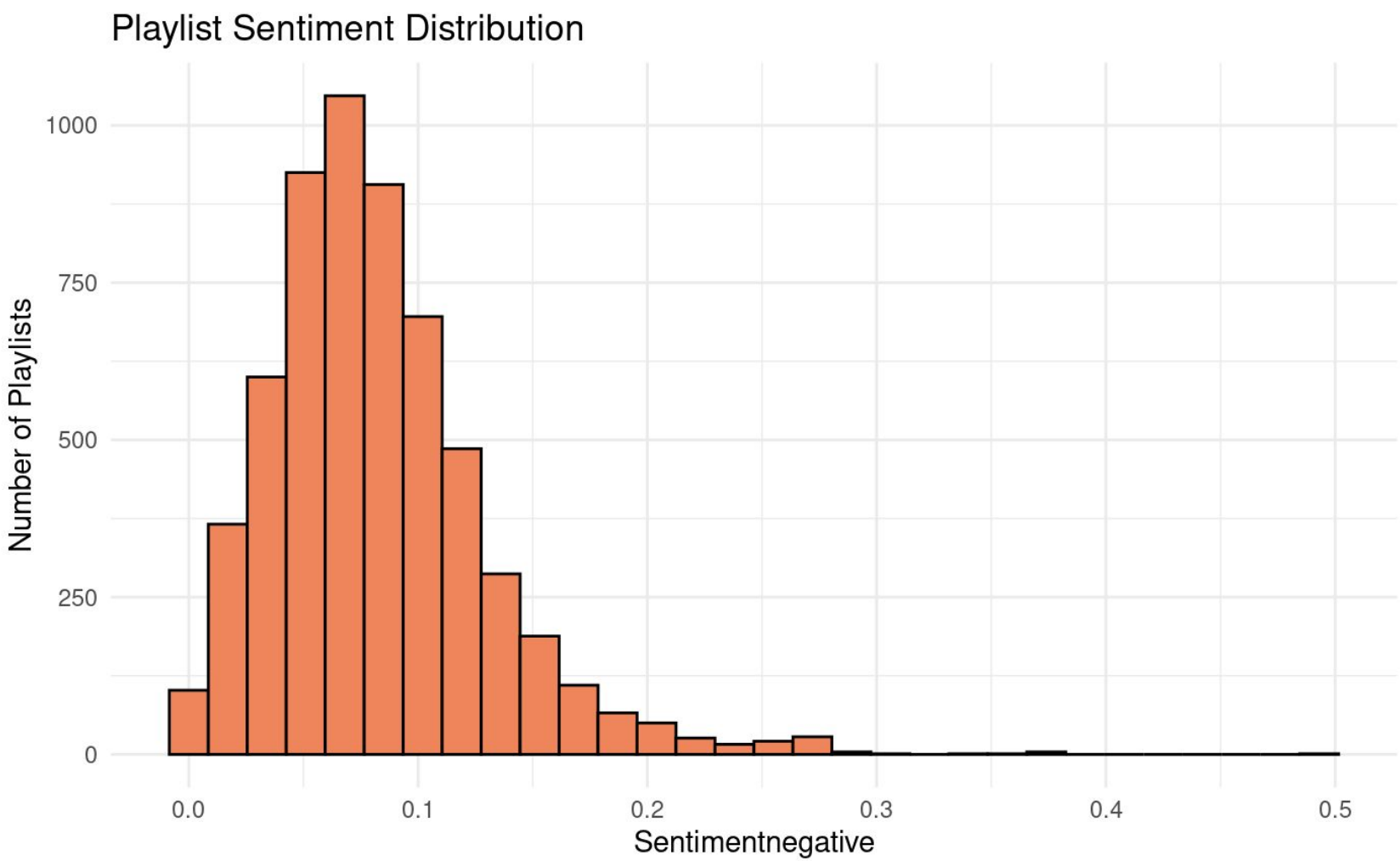
This database contains one million playlists created from 2010-2017 which I will cross reference with SMSD to identify how frequently songs from SMSD appear within the playlists from SMPD.

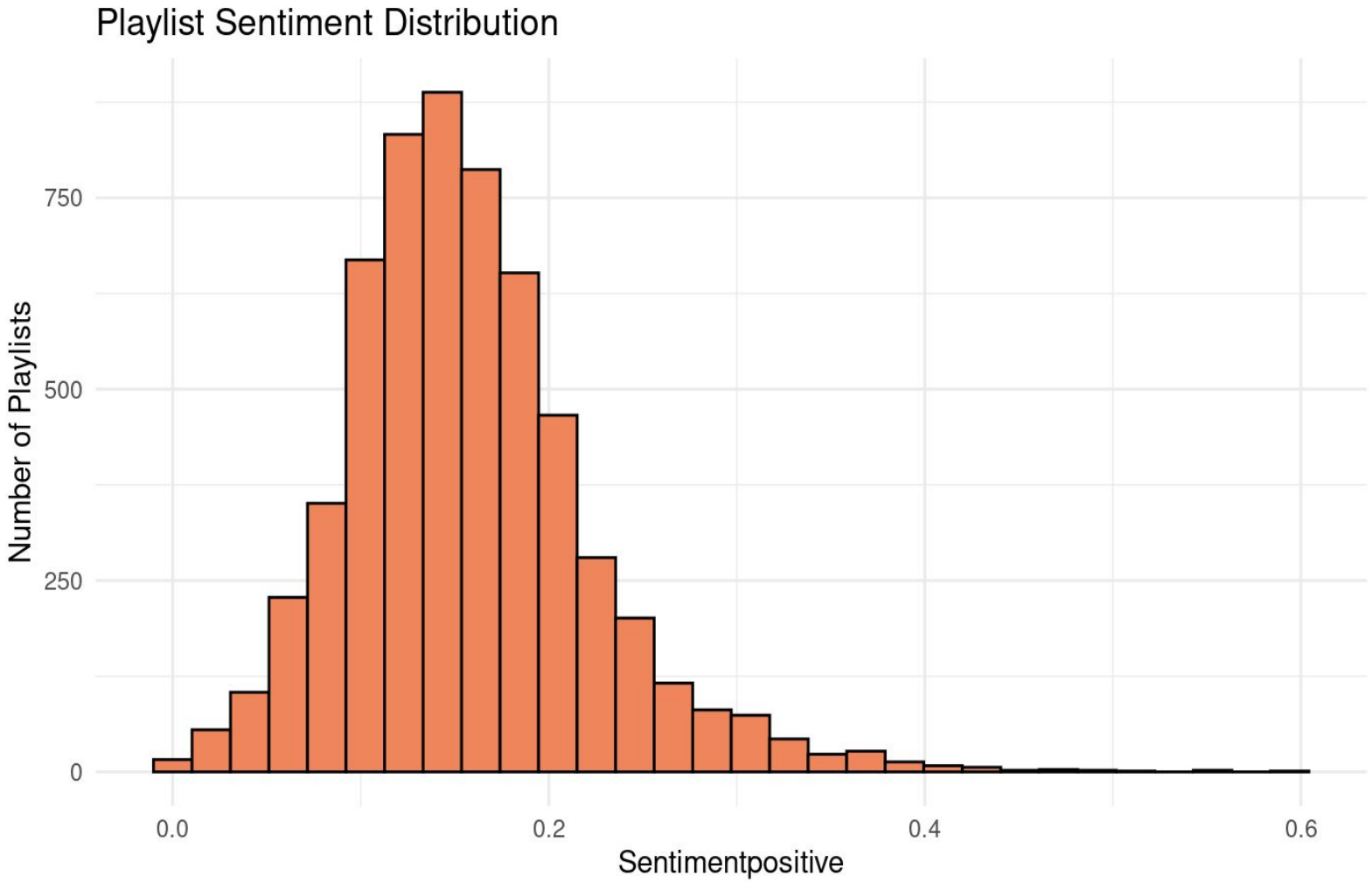
Methodology

- Load and Clean datasets,
- Verify that names and titles are correct across playlist and song datasets
- Identify most frequently occurring songs
- Apply sentiment analysis on lyrics of songs
- Create word clouds of most frequently appearing words
- Create charts showcasing the distribution of positive and negative language in spotify playlists.


Playlist Sentiment Distribution
Number of Playlists
4000
3000
2000
1000
0
Negative
Neutral
Positive
Sentiment


Distribution of Playlist Sentiment Scores
Frequency
2000
1500
1000
500
0
-1.0
-0.5
0.0
0.5
1.0
Average Sentiment (VADER)


Playlist Sentiment Distribution
Number of Playlists
1000
750
500
250
0
0.0
0.1
0.2
0.3
0.4
0.5
Sentimentnegative


Playlist Sentiment Distribution
Number of Playlists
750
500
250
0
0.0
0.2
0.4
0.6
Sentimentpositive

smooth
confident
beautiful
happiness
paradise
awesome
happy
savior
silent
well
lover
darling
envy
heavenly
loved
clean
pretty
pride
joy
redeeming
lean
wonder
glory
finest
tender
love
whoa
calm
pure
hug
strong
pleased
right
grace
shine
faith
free
like
sweet
heaven
good
smile
dawn
thank
glad
peace
merry
rich
holy
better
warm
bright
shiny
wisdom
celebrate
ready
beauty
hot
best
nice
fine
enough
mighty
radiant
daring
joyful
fiery
lovely
euphoria
smiles

Impact of Project

- Able to conclusively say that the average spotify playlist has more positive sentiment
- With respects to language, the songs that seem to be more positive utilize words relating to love and romance
- Gained comfortability working with R and Python, alongside proficiency in the application of sentiment analysis via Vader

Limitations

- Initial dataset was limited in terms of inferring ways to pursue analysis due to a lack of numerical variables
- Both datasets lacked a categorical variable like year or country of origin
- The patience of working with a million songs or a million playlists
- Only able to work with 1% of the second dataset
- Lack of information on first dataset and why its created in that way.

Potential Future Work

- Run the analysis on all of the data to determine if conclusion was correct
- Move into a more specific analysis using lexicon 10 to keep track of more emotions
- Pursue the idea of making a playlist builder based on the emotional experience for user