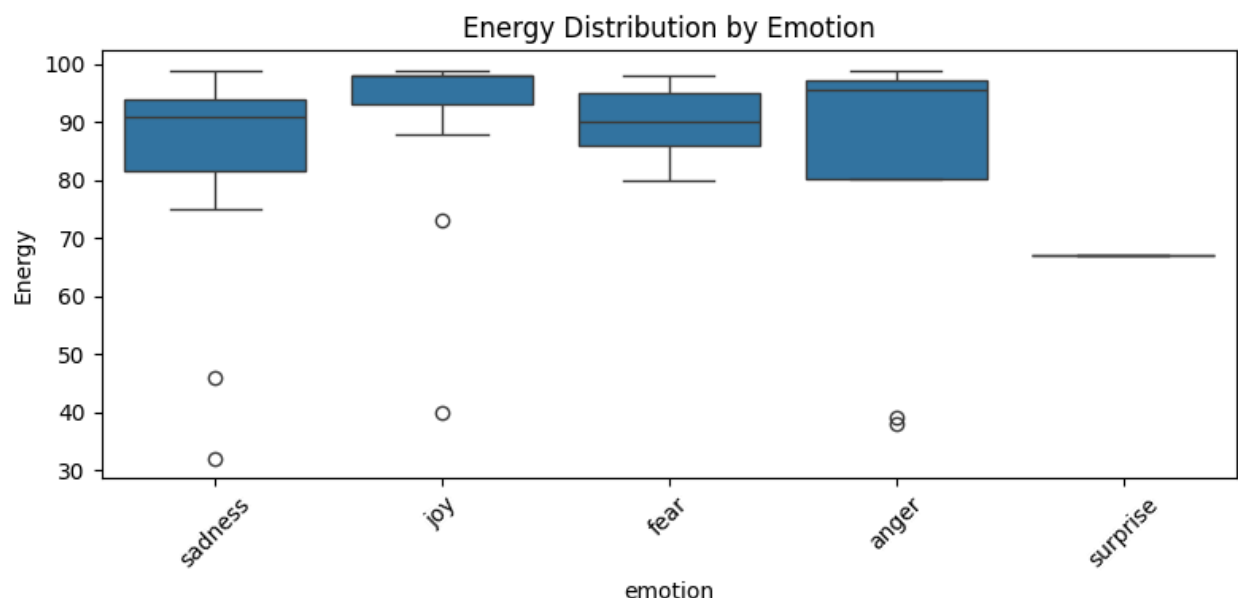# Introduction:

The goal of this project is to explore how musical features from Spotify songs relate to the emotions they convey. For this study we are using a dataset that includes various song attributes along with the emotion of each song. We are trying to figure out which features influence the emotional perception of the song. The goal is to gain a deeper understanding of how songs evoke feelings, and also enable users to pick songs based on their desired emotional impact.
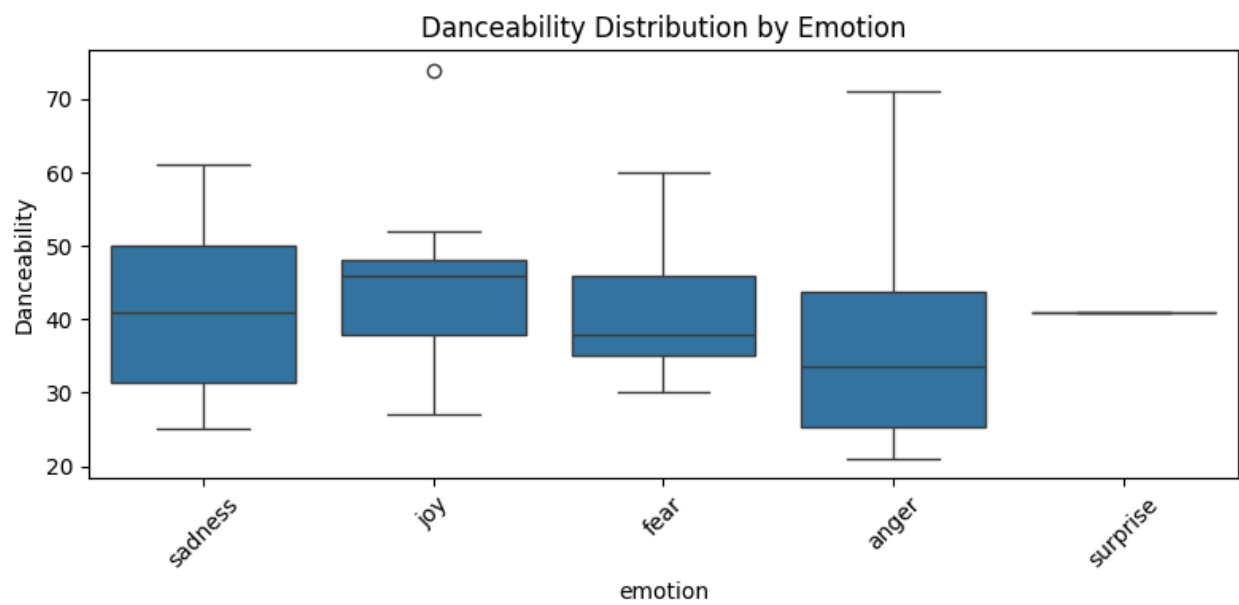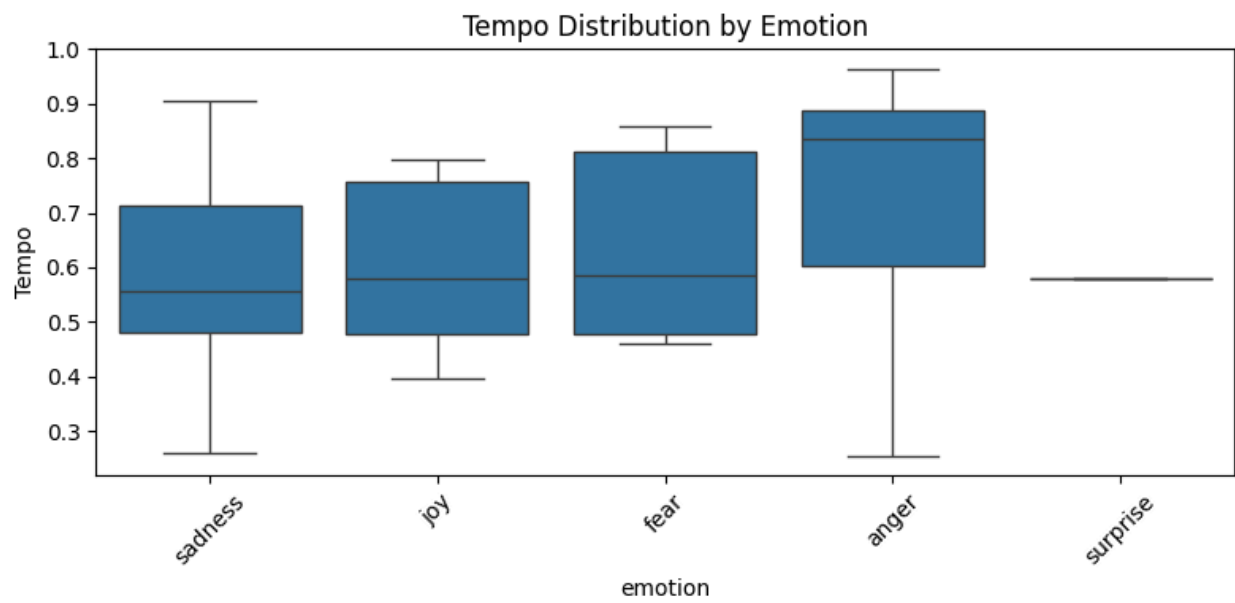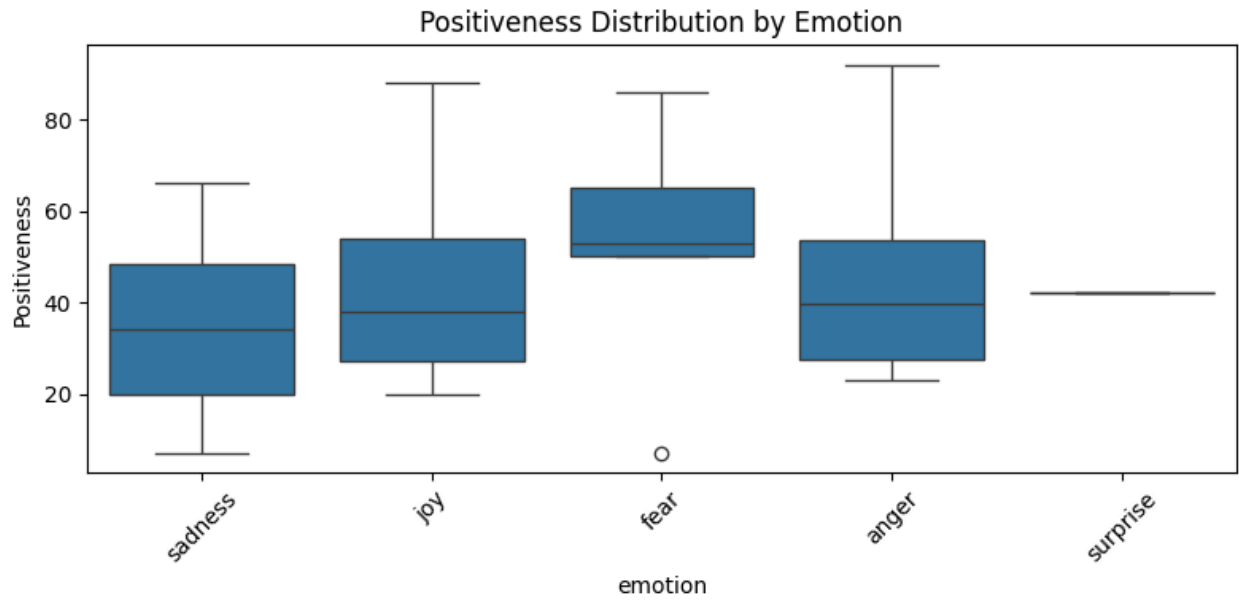
About The Data:

The Data we are using for this study is a CSV file taken from Kaggle user DEVDOPE, it is around 1gb of data.
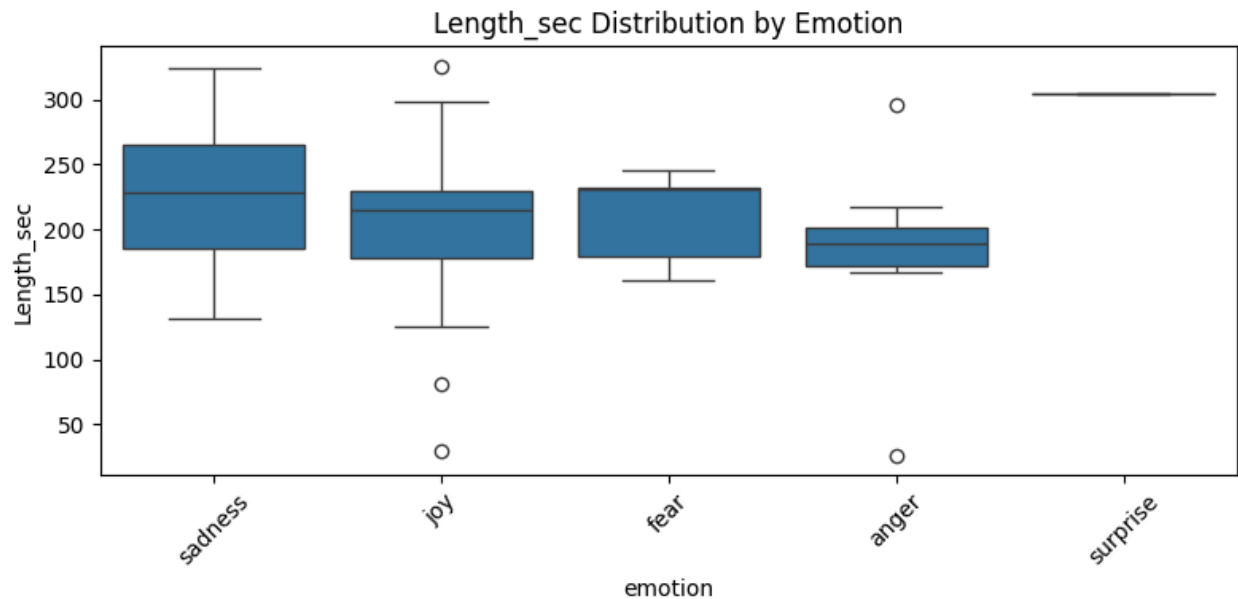https://www.kaggle.com/datasets/devdope/900k-spotify
The data has over 900k songs, and 30 different audio features for every song. Here are some extra visuals I created to try and better understand the data:

Danceability Distribution by Emotion

Positiveness Distribution by Emotion

Tempo Distribution by Emotion

Length_sec Distribution by Emotion

So based on the graphs here we can draw a few inferences: Songs that are joyful have the most energy, sad songs are the most danceable, fearful songs have the most positiveness, songs with anger have the most tempo, and sad songs tend to be longer in length. Some of these don't really make sense, such as sad songs being the most danceable.

Pre-Processing Steps:

First we load the data:
```
df = pd.read_csv("spotify_dataset.csv")
```

Then we drop columns that are not needed:
```
columns_to_drop = [
    'Artist(s)', 'song', 'text', 'Album', 'Release Date',
    'Similar Artist 1', 'Similar Song 1', 'Similarity Score 1',
    'Similar Artist 2', 'Similar Song 2', 'Similarity Score 2',
    'Similar Artist 3', 'Similar Song 3', 'Similarity Score 3',
    'Length'
]
```

Then we drop rows with missing data: `df.dropna(inplace=True)`

## Visualization/Evaluation:

Here is the result of my analysis using classification and random forest :
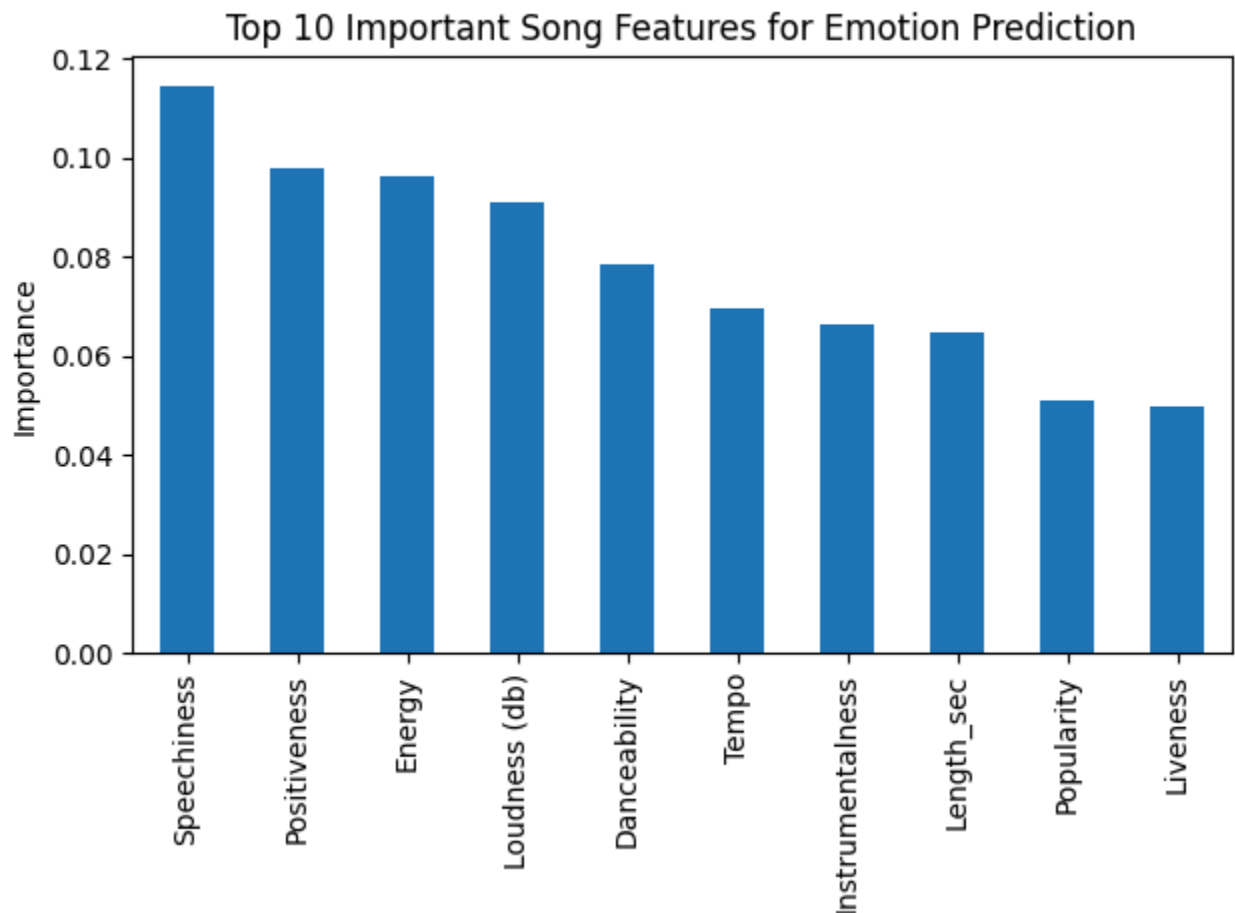


These variables seem to be the most important features when it comes to influencing emotional perception of a song, meaning in our goal to help users choose songs based on emotions, we would likely use these two song features to do that. This helps answer the question asked in the beginning regarding figuring out song factors that influence the emotion.

## Storytelling:

I believe I was able to answer my initial problems in the beginning. The study I conducted gave me useful information regarding song factors that influence emotion. I think to improve my results I would just need extra time to use more data mining techniques to see which one gives the best results, I do believe my initial goal was completed as I now have the top 10 song features that influence emotion, which is pretty much what I wanted from the start.

## Impact Section:

Socially, the impact of this project is that it can be used in real life to help users pick a song based on what kind of emotion that they want to feel. The downside to this is that the models will never be 100% accurate and a user will end up listening to something that they didn't want to listen to. This scenario also applies ethically because a new feature might be advertised to help you find the perfect song but it can never truly work all the time so you will probably end up listening to something that is the opposite of what you wanted and expected.