# Spotify App Review Analysis

**Pochu Sai Prashanthi – U00913293**

Table Of Contents

1. Introduction 2

2. Data preparation 4

- Tokenization 6

3. Text Analysis 8

- Terms 8
- Word Clouds 10
- Topic Modeling 12
- Sentimental Analysis 15

4. Conclusion and Recommendations 19

5. References 22

# Introduction

Spotify is a popular music streaming platform that offers a wide variety of music and audio content to its users. With millions of active users, Spotify receives numerous reviews from its customers daily. These reviews provide valuable insights into the users' experiences with the platform, including their likes, dislikes, and suggestions for improvement. Analyzing these reviews can help Spotify identify the areas where they need to focus on to enhance their users' experience. Therefore, in this project, we aim to analyze the text data of Spotify app reviews using natural language processing techniques to extract insights that can be used to improve the platform's performance.

About Data: This dataset contains reviews of Spotify App from 1/1/2022 - 7/9/2022 collected from Google Play Store.

Total Rows 61594

This dataset consists of the following columns:

Time submitted: The time when the review was submitted by the user.

Review: The text of the review submitted by the user.

Rating: The score given by the user for the reviewed item, on a scale of 1 to 5.

Total_thumbsup: The number of users who found the review helpful and gave it a thumbs up.

Reply: Any response or reply provided by the reviewed item's owner or moderator to the user's review.

Based on the analysis of user reviews of Spotify app using natural language processing techniques, I plan to extract valuable insights to help understand how users perceive the app.By creating word clouds for good and bad reviews, I aim to identify the most frequently used positive and negative words to better understand what users appreciate and dislike about the app. With topic modeling, I intend to identify the areas where the app is lacking and where it is performing well based on ratings. Finally, i aim to perform sentimental analysis on Spotify app reviews to analyze and understand the sentiment expressed by users in their reviews and classify the reviews as positive, negative and gain insights into user satisfaction, identify areas for improvement, and enhance the overall user experience of the Spotify app. This project report includes insights gained, visualizations such as word clouds and topic models, to better understand the patterns and trends in the data which help stakeholders to identify where the app is performing well and the areas where it needs to be improved.

# Data Preparation

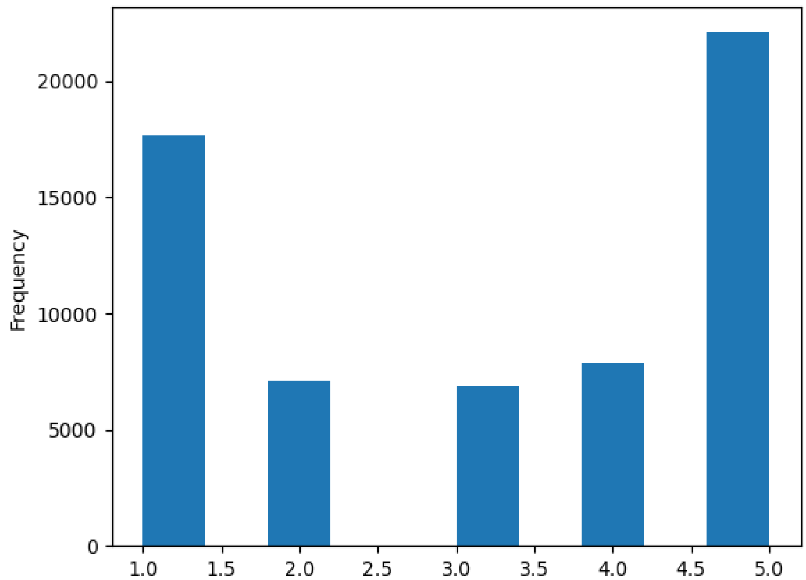
Data preparation is a crucial step in any data analysis project. Data preparation or data preprocessing in general involves not only the transformation of data into a form that can serve as the basis for analysis but also the removal of disturbing noise.

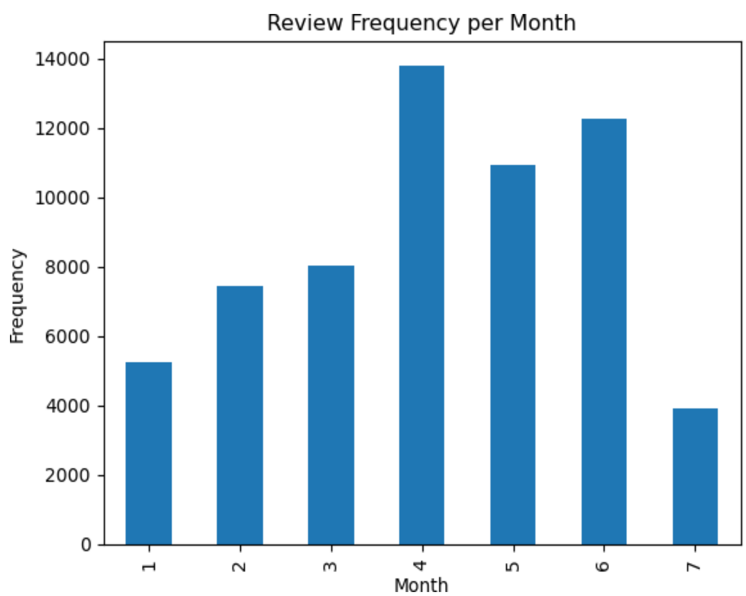
The data preparation phase for Spotify app reviews analysis involves several key tasks. The Spotify App Reviews data which is used in this project is collected from Kaggle. The data contains 61594 rows and 4 columns with 1 column as index namely-Review(The text of the review submitted by the user), Rating(The score given by the user for the reviewed item, on a scale of 1 to 5.), Total thumbs up(The number of users who found the review helpful and gave it a thumbs up), Reply(Any response or reply provided by the reviewed item's owner or moderator to the user's review.); from 1/1/2022 - 7/9/2022.

Initially, the libraries pandas, NumPy, matplotlib, nltk, matplotlib, tqdm, wordcloud and cytoolz are imported into the notebook for data loading, visualization and analysis. Spotify.csv data is loaded into the dataframe df. To know the structure and contents of the data frame df.info() is used which gave following insights:

```
Index: 61594 entries, 2022-07-09 15:00:00 to 2022-01-01 00:19:09
Data columns (total 4 columns):
 #   Column           Non-Null Count   Dtype
---  -------          ---------------- ------
 0   Review           61594 non-null   object
 1   Rating           61594 non-null   int64
 2   Total_thumbsup   61594 non-null   int64
 3   Reply            216 non-null     object
dtypes: int64(2), object(2)
```

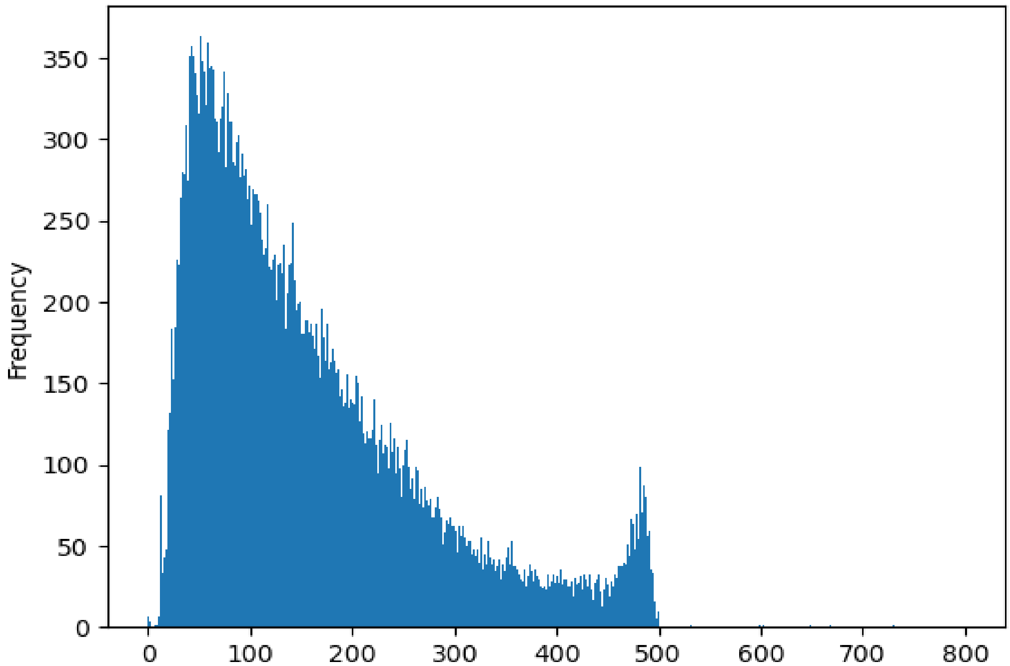
As the Reply column does not have much data and is not significant in this project, this column is dropped from data frame. In the data frame, Time_Submitted is there as an index. As there will be an analysis on the time of reviews this is converted as a column and the data type is changed to date time, which will be helpful for further analysis.




A histogram is drawn to know how the data is distributed with respect to Ratings ranging from 1 to 5. The above bar plot that displays the frequency of reviews submitted per month. Both the graphs shows that the reviews are distributed well among the Ratings and Month. Most of the reviewers are most either satisfied or most dissatisfied with the app. Majority of the reviewers are satisfied with the app.

To perform text cleaning on the Reviews column, which helps in text analysis of reviews, clean text() method is defined, removing any characters which are not alphabets, splits the Review text into individual words, and the cleaned text is stored in a Review column of data frame.




This histogram plot shows the distribution of the lengths of the "Review”. The length of reviews is ranging from 0 to 550, with most reviews in the range of 0 to 200. Overall, the data is prepared and good to go for data analysis.

## Tokenization

For tokenization of Reviews spacy is used, as it provides a robust and efficient tokenization process and handles complex cases like contractions, punctuation, and special characters well, while also providing access to other powerful natural language processing functionalities.

The spaCy Matcher is used to find specific patterns in text. It allowed to define patterns based on linguistic features, such as part-of-speech tags or dependency relationships and match those patterns within the text. This helped in tasks such as identifying specific noun phrases or extracting structured information from unstructured text.

Then, get phrases() function is defined that takes a document doc as input and returns a list of phrases extracted from it. This function is then applied to each review in the Data Frame "df" using spaCy's pipeline, processed in batches with parallel processing for efficiency. The identified phrases are then stored in the candidates list and dumped in candidate_spotify.pkl file for later use.

Then, a defaultdict called freqs is initialized to store frequency distributions of different phrase lengths, which will help in extracting terms from Reviews column. The loop iterates over the candidates list, calculates the length of each phrase, and updates the frequency count for that phrase length in freqs.

# Text Analysis

## Terms

### Extracting Significant terms

Extracting terms from candidates, allows us to filter and focus on specific phrases that are potentially meaningful or significant within the context of the dataset or analysis. These extracted terms can be used for various purposes, such as building domain-specific dictionaries,

To do this C-value function is defined, which is a measure used in computational linguistics to determine the significance of multi-word terms or phrases within a corpus. It calculates the term hood of phrases based on their frequency distribution (freqs) and a threshold value (theta). The function iterates over phrases in descending order of length and calculates the term hood using the C-value formula. Sub terms of each phrase are considered, and if they meet the frequency threshold, their longer counterparts are updated. The function returns a term hood frequency distribution for significant terms with a term hood value above the threshold. Finally, c_value is applied to freqs with a threshold of 50, and the resulting terms are stored in the terms variable.

The most common 15 terms used in the Reviews with their frequencies are as follows:

```
[(('music', 'app'), 2275.125),
 (('gíeat', 'app'), 1670.0),
 (('good', 'app'), 1631.0),
 (('many', 'ads'), 936.0),
 (('new', 'update'), 898.0),
 (('last', 'update'), 694.0),
 (('inteínet', 'connection'), 666.0),
```

```
(('favoíite', 'songs'), 625.0),
(('gíeat', 'music'), 552.5),
(('íandom', 'songs'), 544.0),
(('fíee', 'veísion'), 534.0),
(('otheí', 'apps'), 479.0),
(('íecent', 'update'), 460.0),
(('joe', 'íogan'), 439.0),
(('amazing', 'app'), 435.0)]
```

**Inference:** This depicts that most of the reviewers are feeling that the app is great and good. The terms which are needed to be concentrated here are - many ads, internet connection, updates, free version and other apps, which seems like affecting the App in bad or good way. There are some improvements to be made like reducing the number of ads, updating the app in such a way that it becomes user friendly. Improving the app with respect to free versions, updates help the app in gaining more customers.

And the least 15 terms are as follows:

```
[(('joe', 'íogan', 'expeíience'), 53.8887250245193),
(('fíee', 'useís'), 53.0),
(('fíee', 'useí'), 53.0),
(('ads', 'ads'), 53.0),
(('past', 'few', 'days'), 52.30376252379815),
(('píemium', 'family'), 52.0),
(('favoíite', 'podcasts'), 52.0),
(('pc', 'veísion'), 52.0),
(('bluetooth', 'devices'), 52.0),
(('only', 'complaint'), 52.0),
(('similaí', 'songs'), 51.0),
(('fíee', 'píemium'), 51.0),
(('spotify', 'useí'), 51.0),
(('old', 'spotify'), 51.0),
(('nice', 'music', 'app'), 50.718800023076994)]
```

Inference: Joe Rogan is an American comedian, podcaster, and UFC commentator. Most of the users are talking about his content.

If the content created by Joe Rogan is well organized and managed the app could make the customers more satisfied

All the terms extracted are stored in terms_Spotify.txt file for later use. Tokenization of text is done by recognizing and merging predefined phrases based on the terms defined in the 'terms_Spotify.txt' file. The resulting tokens are then normalized and returned, with the option to replace spaces with a specified separator.

## Word Clouds

In order to make word clouds on the Reviews, the data is divided into two parts good with Rating greater than 3 and bad with Rating less than 3. The reviews with Rating 3 are ignored as they are comparatively less and might confuse the main idea of what good and bad reviews are.

Simply tokenizing the Reviews does not help as their will be many insignificant words which occur frequently likes follows:

```
[('i', 31477), ('the', 25821), ('to', 23423), ('it', 21469), ('and', 19273)]
```

In order to overcome this BigramAssocMeasures metrics Pointwise mutual Information is used, which provides valuable tools for measuring and quantifying the association strength between bigrams and aids in identifying significant word associations, selecting informative features, extracting collocations, and enhancing language modeling tasks. We can either use Pointwise mutual Information or Log Likelihood Ratio metrics. In this project PMI is used.

A cloud function is defined, which shows top 40 frequent words.

The following word cloud shows top 50 words which are used in Reviews with rating less than 3.

Bad reviews (by pmi)


The words like garbage_app, terrible_app are expected in bad reviews. But the words which are not quite common are restricting, unauthorized, lisa, stupid_app.

By looking close at the reviews which contains the following words, the following conclusions are made.

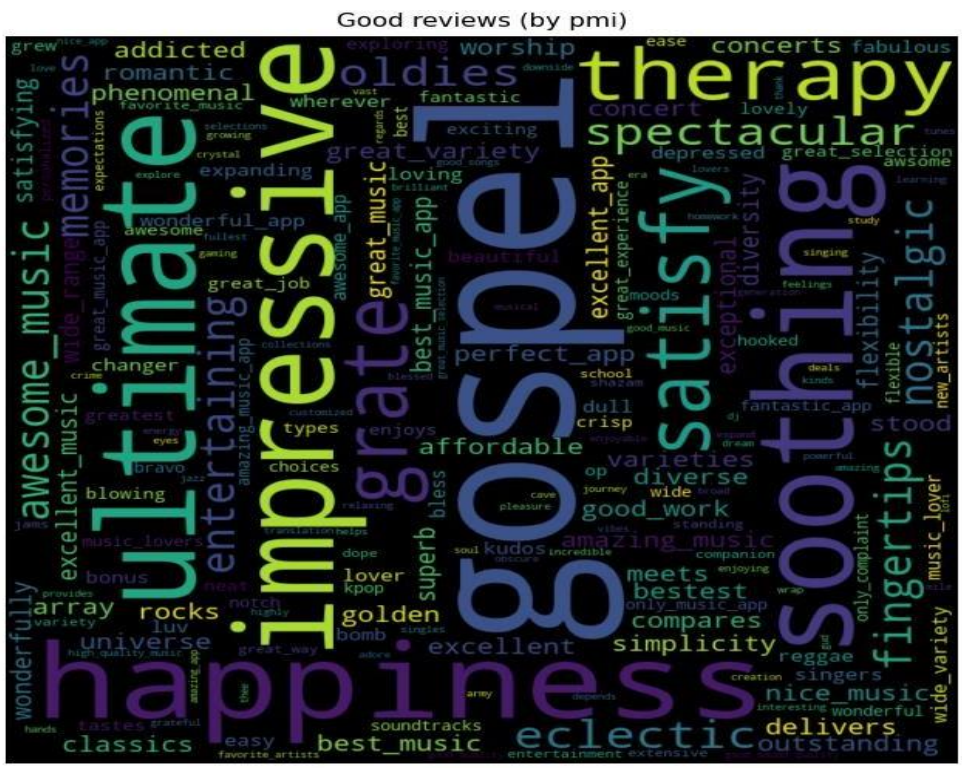
**Inference:**

By analysing the word "stupid app" from reviews, it is clear that people are hating because of its crashing nature.

Lisa, unauthorized: Again it is about the "Money" song by Lisa, which is replaced by an unauthorized song

Seems like Reviewers are restricted from using app as they like. Like playing their favorite song, forcing to buy premium.

The following word cloud shows top 50 words which are used in Reviews with rating greater than 3.




The words like impressive, soothing, happiness are quite common. After looking at reviews containing words like-Gospel, satisfy, finger tips; the following conclusions are made.

**Inference:**

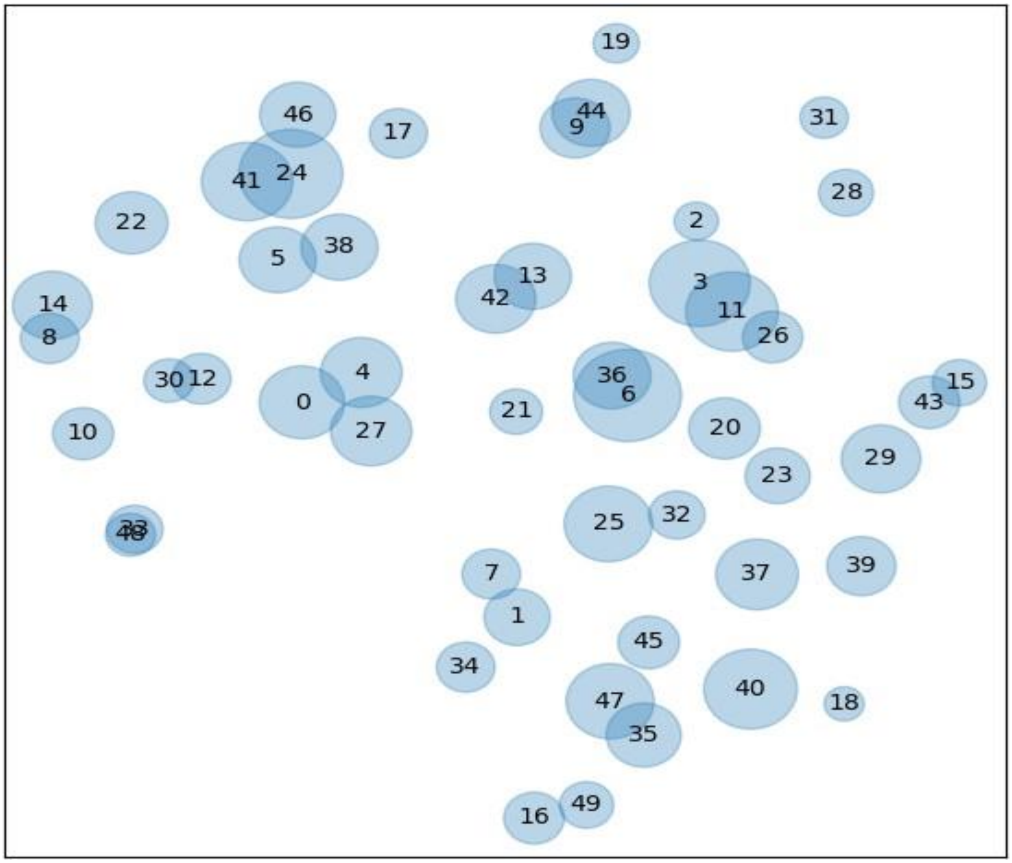
Reviewers are using the word "satisfy" and “gospel” with different contexts like music, app, personal feeling,etc

Reviewers are happy that they can listen to their favourite songs at their finger tips.

## Topic Modeling

Topic modeling can be helpful in analyzing Spotify app reviews by identifying and uncovering the underlying topics or themes present in the reviews.

A popular topic modeling algorithm,LDA stands for Latent Dirichlet Allocation, is used in this project. It is a generative statistical model that allows for the discovery of hidden or latent topics within a collection of documents. It assumes that each document in the corpus is a mixture of various topics, and each topic is characterized by a distribution of words. The goal of LDA is to identify these latent topics and their associated word distributions based on the observed words in the documents.




The above plot shows the top 50 topics for the Reviews. Each cluster represents similar words combined together. Above plot shows that the most common topic numbers are -32,24;39,17,42,30.

The model built is saved in Spotify_topics.bin. The first word of each topic cluster is saved as label and all other terms as words, and saved as Slabels.csv file, whose first three rows are as follows:

| Label | words |
|---|---|
| WORKING. | working, after, fine, then, ago, had, update, again, until, work |
| FIND | find, could, old, here, any, new, time, every, found, everything |
| OVER | over, shuffle, same_songs, same, again, plays, hear, only, repeat, radio |

## Analysis on the Reviews with Rating1 and 5 are made and the following conclusions are drawn:

### Inference on Reviews with Rating 1

And their topic words are as follows:

- ABOUT- about, people, know, their, who, by, money, how, would, because
- STOP- stop, close, pause, randomly, stops, again, sometimes, keeps, force, restart
- UPDATE- update, pause, skip, latest, up, show, stop, bar, controls, player
- PLAYS- plays, then, click, another, try, because, search, album, only, let
- WORKING-working, after, fine, then, ago, had, update, again, until, work
- PHONE- phone, car, bluetooth, connect, devices, works, device, work, google, using
- LISTENING- listening, best, enjoy, favorite, thank, much, thanks, easy, amazing, good_app

Among these the most important topics where users are not satisfied are as follows: STOP and UPDATE which are talking about the controls of the app like, start, pause, stop, restart- there might be something wrong with the controls of app. The app needs to improve its interface.

WORKING- which is talking that the app is not performing well after the updates. This issue should be rectified by the app developers.

PHONE- the topic words of this topic conveys that there is a problem in connecting with other devices. This should be fixed.

## Inference on Reviews with Rating 5

And their topic words are as follows:

- LISTENING- listening, best, enjoy, favorite, thank, much, thanks, easy, amazing, good_app
- ABOUT- about, people, know, their, who, by, money, how, would, because
- EASY- easy, great, podcasts, selection, great_app, excellent, great_music, sound, lots, awesome

- SOME- some, its, great, good_app, sometimes, annoying, pretty, great_app, overall, other
- PLAYLISTS-playlists, find, artists, great, also, make, listening, able, how, new
- EVER- ever, best, best_music_app, best_music, streaming, worst, used, far, out, apps
- YEARS- years, had, using, never, used, since, any, over, problems, always

Among these the most important topics where users are statisfied are as follows:

- EASY: The users are happy with the spotify podcasts. Concentrating more on podcasts might help in attracting more customers.
- PLAYLISTS: The playlists made by the app according to the user previous listened songs worked well. The users are satisfied with this feature.
- EVER: The words in ever indicates that the reviewers are praising the app.

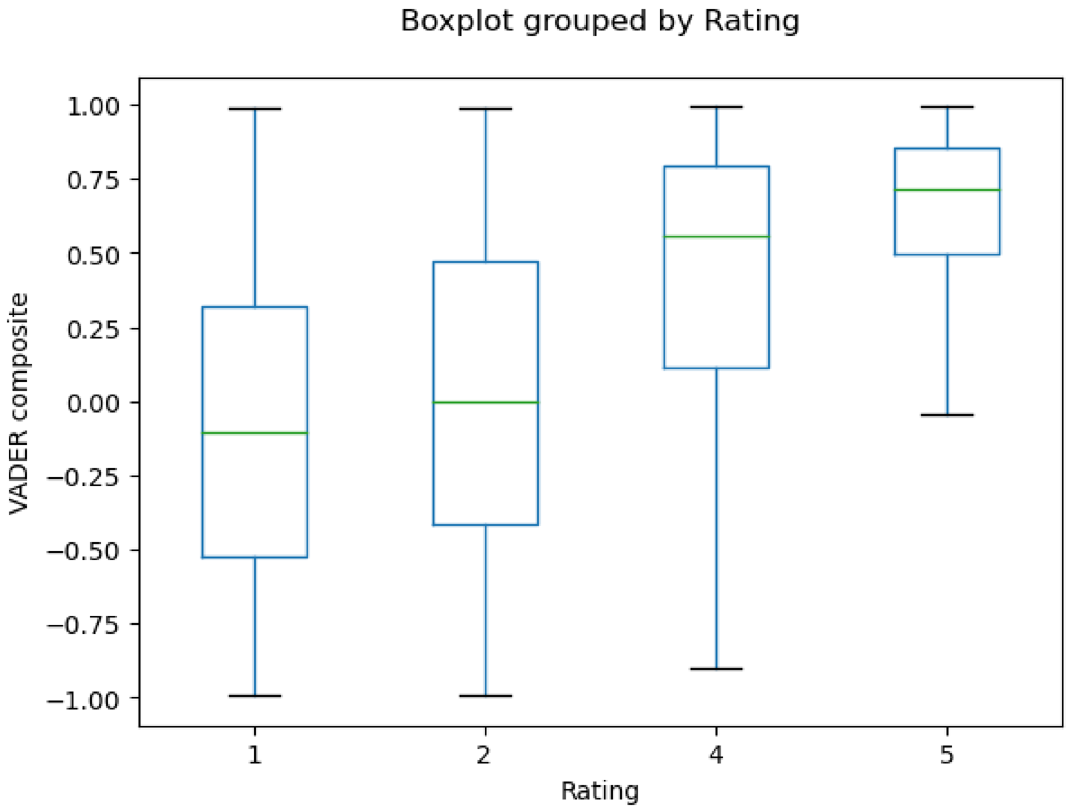
## Sentimental Analysis

Sentiment analysis of reviews on the Spotify app provides valuable insights. By analyzing the sentiments expressed in the reviews, Spotify can gain a deeper understanding of user satisfaction and identify areas for improvement.

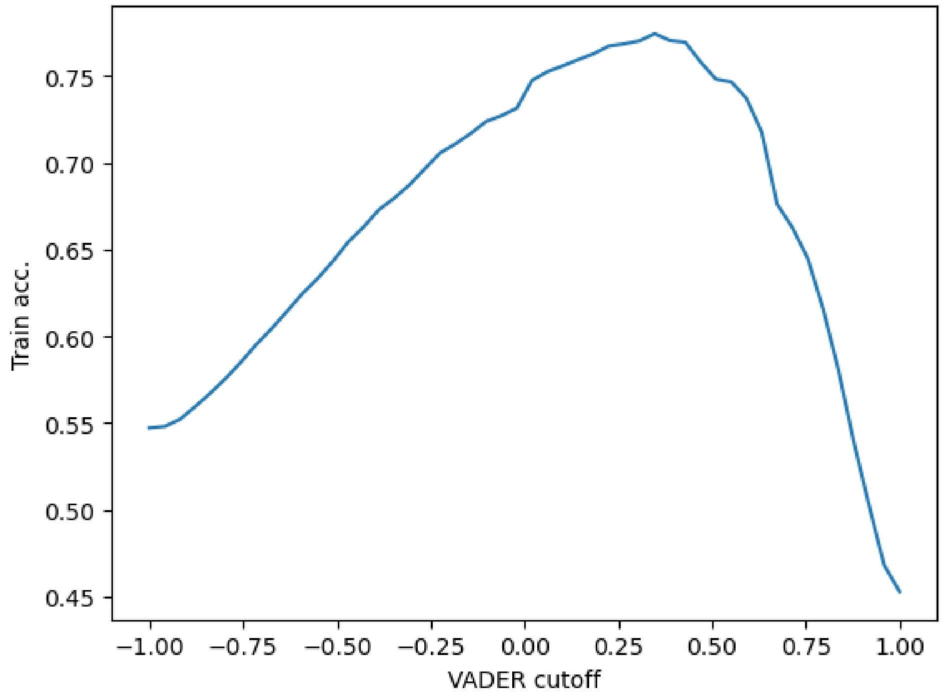
To perform sentimental analysis on the Reviews, the reviews with rating greater than three are considered as positive and the reviews less than three as negative. The reviews with rating three are not considered, as the words used in those reviews would be a mix of both positive and negative words, Thus, confusing the classifiers and other tools which will be used. The Reviews are well distributed with respect to sentiment.

Initially a dummy classifier is used to compare the performance of more advanced classification models. The data is split into 80% training data and 20% test data.This classifier is trained using the "Review" column as the input features and the "Sentiment" column as the target variable from the training dataset. The accuracy of the classifier is then evaluated using the test dataset by comparing the predicted "Sentiment" values with the actual values.DummyClassifier is predicting the correct sentiment approximately 54.72% of the time on the test data.

After this Vader scores for reviews are evaluated.VADER (Valence Aware Dictionary and sentiment Reasoner) is a popular sentiment analysis tool that measures the sentiment of a text using a pre-trained lexicon of words and phrases.




The above box plot shows the distribution of VADER composite scores for different ratings. The plot shows that the reviews with rating 4 and 5 fall in positive sentiment and some of the reviews with rating 1 and 2 also fall on positive sentiment. To eliminate this issue, we can find a cutoff value theta, where compound score greater than theta equals positive review and compound score less than or equal to theta are negative reviews. To do this, we will find a threshold using the training data and then test it on our test data.




When the theta value is 0.3469, it achieved highest accuracy score of 78%.

Syntactically augmented classification in sentiment analysis enhances accuracy by incorporating syntactic information from the text. By considering the grammatical structure, word dependencies, and syntactic patterns, it captures the nuanced sentiment expressed, leading to more accurate sentiment classification and better understanding of sentiment in complex sentence structures. So, syntactically augmented classification is done on the Reviews with negation words.

For the classification, SGD classifier is used,because of it’s effectiveness for sentiment analysis due to its efficiency, scalability, and ability to handle large datasets. It also optimizes the loss function iteratively, making it suitable for online learning. Additionally, SGD works well with sparse data and can handle high-dimensional feature spaces, making it a popular choice for sentiment analysis tasks. Initially to indicate the words in the scope of negation, prefix NOT is augmented to those words and SGD model is fitted to data,which resulted in accuracy score of 89.9%.

Then,the negative marker is spread onto dependent words that come to the right of the negate word. Not all words to the right, though. Just ones that are dependents of the negated word. Again the SGD model is fitted to the data, resulted in the accuracy score of 90%

Next,modifiers are combined with their heads, SGD model is fitted, which resulted in accuracy score of 90.2%.

The most important features (words) that contribute positively or negatively to the classification model's decision-making process, based on their corresponding coefficients by SGD classifier, combining heads with their modifiers are generated. Among them the top 10 with their corresponding coefficients are as follows:

| | | | |
|---|---|---|---|
| love | 6.232 | not | -4.138 |
| great | 5.209 | worst | -3.297 |
| best | 4.517 | misinformation | -2.587 |
| good | 3.971 | horrible | -2.465 |
| amazing | 3.901 | worse | -2.436 |
| easy | 3.230 | bad | -2.411 |
| awesome | 2.976 | terrible | -2.337 |
| thank | 2.652 | sucks | -2.307 |
| nice | 2.467 | now | -2.287 |
| excellent | 2.387 | fix | -2.247 |

# Conclusion and Recommendations

Based on the analysis of Spotify app reviews, several conclusions can be drawn:

**Histograms:** The majority of reviewers either expressed high satisfaction or extreme dissatisfaction with the app. However, the overall majority of reviewers appeared to be satisfied. Additionally, the distribution of reviews was found to be fairly consistent across different months, indicating a steady flow of feedback.

**Term Extraction:** The analysis of extracted terms revealed that users commonly mentioned positive aspects like "music app" and "great app." However, areas for improvement were identified, including concerns about excessive ads, frequent updates, internet connection issues, and limitations of the free version. Addressing these issues could enhance user satisfaction and attract more customers. Furthermore, there was notable discussion about Joe Rogan's content, suggesting that organizing and managing his content could contribute to increased customer satisfaction.

**Word Clouds**: The top negative words mentioned by users highlighted issues such as poor app performance, authorization problems, and frustrating experiences. On the positive side, users expressed satisfaction with terms like "gospel," "happiness," and "impressive." These insights can inform the app's developers about areas that need improvement, such as enhancing performance, resolving authorization issues, and focusing on providing a positive user experience.

**Topic Modeling:** Analysis of reviews based on ratings revealed important topics for unsatisfied users, such as issues with app controls, performance problems after updates, and difficulties connecting with other devices. For satisfied users, positive topics included the ease of using Spotify for podcasts, the effectiveness of personalized playlists, and overall praise for the app.

In conclusion, the analysis of Spotify app reviews suggests that while most users are satisfied, there are areas where improvements can be made to enhance user experience. Addressing concerns related to ads, updates, internet connectivity, and free versions can contribute to greater user satisfaction and attract more customers. Additionally, focusing on optimizing app controls, resolving performance issues, and improving device connectivity can address the concerns of unsatisfied users. Concentrating on podcast offerings and refining personalized playlist features can further enhance user satisfaction.

Based on these results, the following recommendations can be made:

Improve user experience: Address issues related to app controls, performance after updates, and device connectivity to enhance user satisfaction.

Minimize ads and improve free version: Take steps to reduce the number of ads and enhance the free version to provide a more enjoyable experience for users.

Enhance content organization: Ensure that content from Joe Rogan and other creators is well-organized and managed within the app to further increase user satisfaction.

Focus on podcasts and personalized playlists: Capitalize on the positive feedback regarding podcasts and personalized playlists, as these features have garnered user satisfaction. Investing in these areas could attract more customers.

Continuously monitor and respond: Regularly analyze user reviews to stay updated on evolving preferences and address any emerging issues promptly. Engage with users to show responsiveness and commitment to improving their experience.

By implementing these recommendations, Spotify can strive to enhance user satisfaction, address concerns, and attract a larger customer base.

## Conclusion for sentimental Analysis:

Based on the results of the sentiment analysis project on Spotify app reviews, different approaches were applied to enhance the accuracy of the SGD classifier.

Based on these results, it can be concluded that combining heads with their modifiers was the most effective approach for sentiment analysis on Spotify app reviews. It outperformed the other two approaches and achieved the highest accuracy score. This suggests that considering the relationship between words and capturing their modifying context can lead to more accurate sentiment analysis results.

However, it is worth noting that even with the highest accuracy achieved, there is still room for improvement. Further exploration of different techniques, feature engineering, or the use of more advanced models may help enhance the sentiment analysis accuracy even further in future iterations of the project.

# References


- Jens Albrecht, Sidharth Ramachandran, & Christian Winkler. 2020. *Blueprints for Text Analytics Using Python.* O'Reilly. LINK
- Andreas C. Müller & Sarah Guido. 2017. *Introduction to Machine Learning with Python: A Guide for Data Scientists.* O'Reilly. LINK
- https://www.kaggle.com/datasets/mfaaris/spotify-app-reviews-2022?select=reviews.csv