

notebook

July 22, 2021

1 Competing in the Mental Wellness iOS App Market (Natural Language Sentiment Analysis)

Author: Jonathan Lee

GitHub Repository: <https://github.com/ds-leehanjin/app-reviews-nlp-sentiment-analysis>

1.1 TABLE OF CONTENTS

Click to jump to matching Markdown Header.

- **Introduction**
- **OBTAIN**
- **SCRUB**
- **EXPLORE**
- **MODEL**
- **INTERPRET**
- **Conclusions/Recommendations** _____

2 INTRODUCTION

Over the course of the last year and a half, more and more COVID-19 lockdowns and restrictions were put in place, it's of no surprise that the number of people seeking help with anxiety and depression has skyrocketed. According to [Mental Health America](#), there was a 93% increase in the number of anxiety screens and a 62% increase in the number of depression screens from January to September of 2020 in comparison to the total number of anxiety and depression screens recorded in 2019.

To make matters worse, lockdowns have made it increasingly difficult for those people to arrange to meet with health professionals who can provide the assistance that they seek. This is where Mental Health mobile apps have begun to provide an alternative approach. As a result, the mental health app market has seen tremendous growth, and [Market Watch](#) predicts the Mental Health app market to see a compound annual growth rate of 20.5% from 2021 through 2027 to reach \$3.3 billion by 2027.



2.1 Business Problem

Our stakeholder wants to enter the mental health mobile app market and requires insight into what to focus on while designing a new mental health app. With over 300,000 health apps currently available, there is plenty of information available to help us begin to form a strategy.

The goal of this analysis is to use Natural Language Processing to determine what characteristics of Mental Health apps currently available on the iTunes App Store are liked or disliked by users. Through sentiment analysis of text reviews and their corresponding ratings, we will examine what words are more likely to indicate a positive vs a negative app review.

Questions to consider: * What do users like or dislike about apps currently available on the iTunes App Store? * How can we use this information to develop a strategy for building a new mental health app that can compete with apps that have already seen success? ***

3 OBTAIN

3.1 Data Understanding

Although there is a category for “Health and Fitness” apps, there is no subcategory for mental health apps, and so a list of 31 top mental health apps was hand-curated from a collection of blog posts. The data used in this analysis was scraped using the [itunes_app_scraper](#) and [app_store_scraper](#) libraries and includes text reviews for this list of mental-health apps paired with ratings out of 5. The process for scraping the data can be found [here](#).

The review texts will be our independent variables, and our target variable will be the rating out of 5.

3.1.1 Installing and Importing Libraries

Let’s begin by installing and importing the necessary libraries to run this notebook as well as changing some settings to be able to easily visualize our data.

```
[1]: # Install LIME library for feature interpretation
!pip install lime
```

```
Requirement already satisfied: lime in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (0.2.0.1)
Requirement already satisfied: numpy in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from lime) (1.18.5)
Requirement already satisfied: scipy in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from lime) (1.5.2)
Requirement already satisfied: tqdm in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from lime) (4.50.2)
Requirement already satisfied: scikit-learn>=0.18 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from lime) (0.23.2)
Requirement already satisfied: matplotlib in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from lime) (3.3.1)
Requirement already satisfied: scikit-image>=0.12 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from lime) (0.18.2)
Requirement already satisfied: joblib>=0.11 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from scikit-learn>=0.18->lime) (0.17.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages (from scikit-
learn>=0.18->lime) (2.1.0)
Requirement already satisfied: certifi>=2020.06.20 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from matplotlib->lime) (2021.5.30)
Requirement already satisfied: python-dateutil>=2.1 in
/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages (from
matplotlib->lime) (2.8.1)
Requirement already satisfied: cycler>=0.10 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from matplotlib->lime) (0.10.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from matplotlib->lime) (1.2.0)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.3 in
/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages (from
matplotlib->lime) (2.4.7)
Requirement already satisfied: pillow>=6.2.0 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from matplotlib->lime) (7.2.0)
Requirement already satisfied: imageio>=2.3.0 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from scikit-image>=0.12->lime) (2.9.0)
Requirement already satisfied: networkx>=2.0 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from scikit-image>=0.12->lime) (2.5)
Requirement already satisfied: tifffile>=2019.7.26 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from scikit-image>=0.12->lime) (2021.7.2)
Requirement already satisfied: PyWavelets>=1.1.1 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from scikit-image>=0.12->lime) (1.1.1)
Requirement already satisfied: six>=1.5 in /opt/anaconda3/envs/learn-
env/lib/python3.8/site-packages (from python-dateutil>=2.1->matplotlib->lime)
(1.15.0)
```

Requirement already satisfied: decorator>=4.3.0 in /opt/anaconda3/envs/learn-env/lib/python3.8/site-packages (from networkx>=2.0->scikit-image>=0.12->lime) (4.4.2)

```
[2]: # Import standard packages to be used
import sys
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os
import glob
import string
from PIL import Image
import urllib.request

# Import standard packages for Natural Language Processing
import nltk
from nltk import FreqDist, word_tokenize, regexp_tokenize, RegexpTokenizer
from nltk.corpus import stopwords
from nltk.stem.wordnet import WordNetLemmatizer
from wordcloud import WordCloud

# Import SciKit Learn packages for modeling and model evaluation
from sklearn.linear_model import LogisticRegression, LogisticRegressionCV
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.svm import LinearSVC
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.pipeline import Pipeline, make_pipeline
from sklearn.feature_extraction.text import TfidfTransformer, TfidfVectorizer
from sklearn.inspection import permutation_importance
from sklearn import metrics
from sklearn.metrics import accuracy_score, precision_score
from sklearn.metrics import classification_report, plot_roc_curve
from sklearn.metrics import plot_confusion_matrix, roc_curve

# Import LIME packages for feature interpretation
import lime
from lime import lime_text
from lime.lime_text import LimeTextExplainer

# Magic function to display plots in notebook
%matplotlib inline
```

```
[3]: # Set theme and style for plots.
sns.set_theme('talk')
```

```
sns.set_style('darkgrid')

# Set random seed
np.random.seed(27)

# Set no limit to column width to view full review text
pd.options.display.max_colwidth = None
```

3.1.2 Importing Data

We can now begin by importing data scraped in the data collection notebook. The collected .csv files have already been stored on the project repository, and we can go ahead and use a list of URLs to load them into this notebook.

```
[4]: all_files_github = [
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/headspace-meditation-sleep.
    ↳ csv',
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/calm.csv',
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/stoic.csv',
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/happify-for-stress-worry.csv',
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/shine-calm-anxiety-stress.
    ↳ csv',
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/mindshift-cbt-anxiety-relief.
    ↳ csv',
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/moodtools-depression-aid.csv',
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/insight-timer-meditation-app.
    ↳ csv',
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/moodfit.csv',
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/talkspace-therapy-counseling.
    ↳ csv',
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/moodmission.csv',
    'https://raw.githubusercontent.com/ds-leehanjin/
    ↳ app-reviews-nlp-sentiment-analysis/master/data/sanvello-anxiety-depression.
    ↳ csv',
```

```

'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/minddoc-your-companion.csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/smiling-mind.csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/cbt-i-coach.csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/youper-self-guided-therapy.
↳csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/innerhour-self-care-therapy.
↳csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/breathe2relax.csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/rootd-panic-attack-relief.
↳csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/
↳ten-percent-happier-meditation.csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/pzizz-sleep-nap-focus.csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/noisli.csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/slumber-fall-asleep-insomnia.
↳csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/whats-up-a-mental-health-app.
↳csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/breethe-meditation-sleep.csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/meditopia-meditation-breathe.
↳csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/betterme-calm-sleep-meditate.
↳csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/mindfulness-with-petit-bambou.
↳csv',
'https://raw.githubusercontent.com/ds-leehanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/
↳fabulous-daily-routine-planner.csv',

```

```
'https://raw.githubusercontent.com/ds-leeahanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/reflectly-journal-ai-diary.
↳csv',
'https://raw.githubusercontent.com/ds-leeahanjin/
↳app-reviews-nlp-sentiment-analysis/master/data/relax-meditation-guided-mind.
↳csv']
```

```
[5]: # Create a list of all loaded review text dataframes
df_list = []

for filename in all_files_github:
    temp_df = pd.read_csv(filename)
    df_list.append(temp_df)
```

```
[6]: # Concatenate all tables into a single dataframe
df = pd.concat(df_list, axis=0, ignore_index=True)
df
```

```
[6]:
```

	title	isEdited	\
0	This is the BEST app by far for mindfulness	False	
1	Facing down the rabbit hole	False	
2	A Bright Spot in a Dark World	False	
3	The most useful thing I've done in 2020	False	
4	It's a Miracle App!	False	
...	
44717	Me encanta lo mejor para meditar	False	
44718	Excelente	False	
44719	Love this app	False	
44720	Disappointed	False	
44721	FALSE ADVERTISEMENT subscription terms	False	

	userName	date	\
0	Crazy Yorkie Lady	2021-02-22 18:13:54	
1	KLC-MHFL	2021-02-12 12:42:11	
2	Ashabashley	2021-02-11 06:22:38	
3	Kindly38	2021-01-18 17:14:21	
4	jlhuggins	2021-02-08 19:19:10	
...	
44717	JLucret	2020-04-06 23:34:12	
44718	mata rato	2020-03-28 06:01:54	
44719	Fierce Girl	2020-03-24 13:44:13	
44720	Irkesome	2020-03-07 17:29:09	
44721	ksavv916	2020-02-18 02:08:56	

```
review \
0    It's almost embarrassing to say this, but I have been a "meditator" for
about 30 years and I have to say that I never really got very much out of doing
```

it. I tried so hard, but I'm finding out I never truly understood that "trying" to be a good meditator or mindful person, was missing what doing this was about. What I'm learning through this app is that "mindfulness" is about me - understanding myself first, and it's a lot about acceptance. So with that understanding and acceptance I can begin to see "things" so much more clearly and why previous ways of thinking, particularly about myself, were holding me back from living my life and loving myself. And I'm finding that the more I practice self acceptance, the fears about living fully and going after what I want out of my life, are disappearing. \n\nThe previous version of myself is rapidly becoming someone I only used to know, and this new ME is much more in line with who I'd always wanted to be. I'm able to try and do so much more, all while in the comfort that I am lovable, even if I fail at something. I simply dust myself off and try some more.\n\nThis app has completely changed me and the way I'm able to appreciate all the good that's in my life. And I'm worthy of going after the things I want for my life, while freeing me up to share my gifts with others and society in general.\n\nIf this is where you find your life stalling out, this app will bring you back to yourself.

1 I discovered Headspace on Netflix. My life is a series of invitations to succumb to fear. My husband is a practicing psychologist who is on peritoneal dialysis for 10 hours every night and works remotely with patients for about 10 hours every day without complaint. We have been waiting for a kidney for two years. I am optimistic by nature but even if this was the only focus for concern it would be a lot. Unhelpfully, the list goes on with equivalent issues. I can't work; I need to be on call because I am the go-to for support. I occupy my mind and time writing a book and renovating every room of our home one at a time amidst being available for nine blended (mostly grown) children and 5 grandkids, many of whom have special needs. I am busy but isolated like so many due to covid-19. Trying the Headspace series felt like meeting air; filling the natural and effortless need to breathe. It has provided me with a new lens through which to focus on challenges and see potential. I incorporate it into my waking routine at dawn and it positions me for whatever happens throughout the day, and differentiates a space that is my own. I turn 61 this month and for maybe the first time in a while, life feels lighter and brighter. It's not about what you fear or can't control or even the fact of fear. It's about the lens through which you interpret your fears or challenges and making them your strengths and allies. Thank you Andy :-)

2 You won't regret buying a Headspace subscription. I often joke with people that Headspace is the one thing I would want to take with me to a desert island. If I were to cancel all of my subscriptions today, Headspace is the one exception I would make. I'm a high school English teacher, and our administration put together a month-long mindfulness education program for our students shortly after I subscribed. I had already raved to many of them about the app and the benefits of taking a couple of minutes out of their day to sit and be present. Lo and behold, Andy pops up on the next slide as our Ted Talk speaker of the day! On a more serious note, I suffer from debilitating chronic pain. I'm currently on long-term disability and a leave of absence from the classroom. Mindfulness, and the courses on managing pain, have been a tremendous

help in my journey. Not only that, but I struggle with insomnia. My favorite part of Headspace, because of my (admittedly) vivid imagination, are the sleepcasts. I love falling asleep to Rainday Antiques. I can't imagine doing anything but that ever again. I have learned to manage stress and anxiety in a way that I had never thought was possible. I have meditated in the waiting room of The Cleveland Clinic before my appointments. I once laughed at the idea of mindfulness and meditation. Now it's an integral part of my life. It's part of my journey, and I think everyone could benefit from it.

3

This app has given me tools to deal with constant change. This past year, I've needed to pivot every day to face some new challenge outside as an essential worker, and at home in isolation with my partner. The tools I learned to use to be present in the moment and to accept change has benefited my partner and colleagues at work as much as my own peace of mind. I have very little control over my environment, but I can see my reaction, accept the emotions that rise up, and move on instead of getting lost or bound by them. \nThis app has also benefited me by helping me get through insomnia. Sometimes it helps me get to sleep in the first place. Mostly it gives me tools to use to calm my mind and body when I wake up in the night buffeted by anxiety or lingering frustration from the day. \nLastly, this app has given me a language to express the emotional upheaval and troubling questions that I used to just call anxiety. Naming what makes me feel upset has brought clarity to what can feel like a tempest at the time and allow me to look at my thoughts or, rather, face them. After acknowledging them, I can let them go. Sometimes I'm better at that than at other times. \nRegular exercise strengthens my focus, my ability to pivot and refocus, and builds roads of calm and quiet that the brain finds easier to fall into and travel due to constant use.

4

This app has helped me so much! When I am suffering from stress, health anxiety, or fibro/stress related body pains I turn to this app and go through some meditations with the teachers. It's an excellent way for me to relax. Calm my mind. Stop thinking about myself. And to recover in the moment. Hopefully in time this will help me heal, as previously I've just been getting worse and worse without any therapy since my therapist's practice permanently shutdown from COVID. This has been a great help and is something I can just pull out of my pocket and sit in my bedroom for a bit to calm down. Well worth the 1 year subscription price I've paid. Thank you! \n\n*There is only one thing that I'd change. This would be for some of the workout audios. I need to walk a lot for a heart condition I have (which is a big cause of my health anxiety), and I'd love to see the audios for walks and runs to be a bit longer. The longest run audio I see is 20 minutes, when recommended cardio workout lengths realistically should be at least 30 minutes. I'd really like to have an audio similar to the bedtime audios that are lengthy, have soothing talk, and tips to reduce anxiety made to be listened while walking. And if there could be a variety to choose from. Overall this would be the only change I can think of. An hour long walk audio to help calm my mind.

...

```

...
44717
Es la mejor aplicación se las recomiendo.
44718
Lo recomiendo
44719
Great app.-would love it if the sounds were organized in a way that makes sense.
44720
You get exactly two meditations before you pay 60.00 per year.
44721
As stated in the App Information section WORD FOR WORD: "RelaxMeditation also
offers a lifetime subscription for $19.99 which is paid for by a one-off upfront
payment with unlimited access to RelaxMeditation forever."
My message to App
Support:
"I'm trying to purchase the $19.99 lifetime / unlimited subscription
but can't find where / how to?"
Their response:
"Hello Kristen, thank you
for reaching out to us! I really apologize about this. The offers and prices
displayed in the Store are not accurate, as they display all of our past and
current offers for all of our apps. [...] We do have a special right now, and
it offers 60% off the regular price, for 99.99 USD. [...] We also offer a yearly
subscription, for 59.99$, and it includes a 7-day free trial period[..]"
No 60% off- at $99.
No thanks. Good luck with that.

```

	rating	app_name	app_id	developerResponse
0	5	headspace-meditation-sleep	493145008	NaN
1	5	headspace-meditation-sleep	493145008	NaN
2	5	headspace-meditation-sleep	493145008	NaN
3	5	headspace-meditation-sleep	493145008	NaN
4	5	headspace-meditation-sleep	493145008	NaN
...
44717	5	relax-meditation-guided-mind	367506176	NaN
44718	5	relax-meditation-guided-mind	367506176	NaN
44719	4	relax-meditation-guided-mind	367506176	NaN
44720	2	relax-meditation-guided-mind	367506176	NaN
44721	1	relax-meditation-guided-mind	367506176	NaN

[44722 rows x 9 columns]

4 SCRUB

4.1 Data Preparation

Prior to preprocessing our text for Exploratory Data Analysis, we will check for null values, duplicated rows, as well as any other issues with the text reviews that we might want to address. Note that we have almost 45K reviews to work with.

```
[7]: # Check for null values and data types
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 44722 entries, 0 to 44721
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   title                  44722 non-null  object
1   isEdited                44722 non-null  bool
2   userName                44722 non-null  object
3   date                   44722 non-null  object
4   review                 44721 non-null  object
5   rating                 44722 non-null  int64
6   app_name                44722 non-null  object
7   app_id                 44722 non-null  int64
8   developerResponse       4765 non-null   object
dtypes: bool(1), int64(2), object(6)
memory usage: 2.8+ MB
```

```
[8]: # Check value counts for isEdited column
df['isEdited'].value_counts()
```

```
[8]: False    44425
      True      297
      Name: isEdited, dtype: int64
```

```
[9]: # Check value counts for developerResponse column
(df['developerResponse'].isna()).value_counts()
```

```
[9]: True      39957
      False    4765
      Name: developerResponse, dtype: int64
```

```
[10]: # Preview sample developer responses
df[df['developerResponse'].isna() == False]
```

```
[10]:
```

	title	isEdited	\
29	Loved at first	False	
53	Amazing but missing something	True	
58	Doesn't work well with Iwatch	False	
93	Great, unless you have an issue	False	
98	Used to be better	False	
...	
44571	Amazing choices for sounds, guided meditations okay	True	
44575	Good... but also bad	False	
44603	Time to delete	False	
44618	Can't use it	False	
44624	Great so far	True	

	userName	date \
29	nizismom	2020-04-28 00:59:49
53	jellyfishqueen	2021-02-24 21:41:55
58	njalexsmith	2020-08-09 12:42:14
93	BethicaB	2021-05-15 05:37:35
98	Patricia Lag	2020-10-12 18:22:41
...
44571	F.Ash1088	2020-05-04 21:39:15
44575	kat132904	2020-01-18 22:19:54
44603	bahsgrl	2020-01-09 07:08:51
44618	Cold and black	2020-01-14 16:48:13
44624	J4m3\$rawdogger	2020-05-26 19:55:30

review \

29

I loved it the first day I tried it in February 2020 and purchased a year long subscription. I use it several times a day and my four year old and I fall asleep to the sleep casts every night. I cope with severe anxiety on a daily basis and during the pandemic this has been extremely challenging. Three nights ago I really needed my favorite windown and sleepcast and they are all locked and it's asking me to start a free trial if I want to use them. I have followed all the advice on the help section to try to restore my purchase but even under my subscriptions it shows I am a subscriber and it doesn't expire until feb 2021. I've emailed headspace and have heard nothing back and also tried the chat and it just feeds me the same unhelpful help and then prompts me\nto email them again. If I get this resolved I'll update and add more stars but I'm extremely disappointed right now and not happy with the lack of customer service. Maybe I will switch to Calm

53

This app is a game changer and is leaps and bounds above anything this accessible to the public that I have seen before. They fixed many bugs I once had issues with. The sleep casts no longer have issues buffering\n\nOne key difficulty I have with the app is that it leaves a perspective of naïveté. This app will teach you about tools for everything from meditation to career and money management, but it will not teach you nor properly prepare you for genuine malevolence that likely inevitably arises. \n\nI would caution people to not let mindfulness techniques promoting optimistic mentalities obscure an important skill for developing real danger. The animations can mislead you into thinking mindfulness and being a healthy human goes hand in hand with being naive. This may be more of a problem for some than others, but I felt it helpful to express anyhow since it's not discussed.

58 I wish this app was designed to be more intuitive with the Apple Watch. I can only play a 1 minute breathing exercise on the iwatch app despite the website is conveying and customer support is poor. I sent an email to customer support. The first response from customer service is that my response will take longer than 48 hours with an apology. Second response was a name of a customer service agent and never heard back. I tried reaching out by DM on social media,

but no response. Furthermore, if you use the 1 minute breathing exercise on the watch, it doesn't sync to your account. \n\nThe phone app is great, but I would like to use my phone less, not more. My phone is integrated with my work apps (email, calendar, teams, etc) and adds endless anxiety. Because the phone app is good, I wouldn't give it just one star, however I am posting a 1 star review because both devs and customers only read 1 star reviews. Headspace has a customer service problem along with a poor iwatch app. I haven't had an issue with the phone app.

93

After using this for a few days, I knew it was going to help me and was so excited. I shared the info with my family and a couple of them were interested so I decided to upgrade to the family plan. This is where everything went off the rails. Four emails later and I still have no resolution. I couldn't "upgrade", I was told that I had to cancel and could then upgrade. I canceled but still couldn't upgrade. Then I was informed that I needed to go through Apple to request a refund and I did but I still can't upgrade. Now I'm being told I can't upgrade until the app basically decides I'm no longer a member (and basically doesn't work) in order to upgrade. All my meditation tracked progress is gone but the app still works. However it isn't really relaxing me with all of this frustration and no real help. I've wasted more than two weeks on trying to fix something that shouldn't be this difficult. Very disappointed.

98

I've recommended Headspace to countless people over the years but am disappointed with all the recent updates. Missing the days when Headspace was a clean and simple space that didn't try to be anything other than what it was (a meditative respite). I look forward to the mindful moment notification every day, but now I keep on getting push notifications for working out within the app or to watch a new video with some influencer I couldn't care less about. The only notification I have turned on is for mindful moments and I still keep on getting these other notifications even though there's nothing mindful about them. Please Headspace, stop trying to be like every other app with your incessant notifications and suggestive design, and go back to being the app we all loved you for in the first place.

...

...

44571

Edit: the restore button worked so I did get my lifetime subscription back. I am bringing it up to 4 stars, because it is an excellent program except for the guided meditations being so short and the narrators speaking too quickly. \n\nSo many sounds to choose from to help soothe you that you can try a different one or different combination each day. I also appreciate the isochronic tones, which kept me coming back almost every day. The meditations are good but short. The guides of the meditations speak very quickly and there is no option to slow their speech-if you're trying to meditate, it defeats the purpose to have to keep stopping and going back because they're talking too fast. But I bought a "lifetime" subscription but I have lost almost all of the sounds and meditation options I had because I don't have "pro" despite the fact that they have made no

changes or improvements since I bought it.

44575

I've had this app for quite some time now, and I have different mixes for each day of the week and I use one each night. It works very well and rarely glitches, but many things, you have to purchase. It's annoying. You have to buy sooooo many of the sounds, meditations, and pretty much everything. The app would be much better if almost everything didn't have to be bought. I would appreciate it if this issue was fixed. Thanks

44603

I've been using this app for awhile, and I absolutely loved it. A mix of isochronic tones and a gentle breeze through a grassy meadow was enough to send me off to sleep within minutes of my head hitting the pillow. However, with this latest update, they want me to pay almost 50 bucks for something I've been enjoying for free...I'm calling shenanigans!

44618

I had purchased lifetime access. With last update I was locked out I have contacted developer multiple times I have followed their instructions on how to get access back... now no issues is I can't actually down load ANY meditations without internet access. DOESN'T work over cellular service!!!!

44624

Really liking the app so far. Would like more variety of meditations and sounds though. FYI This app advertises a lifetime membership for \$20 and they don't actually offer that deal, so that is extremely misleading!

	rating	app_name	app_id \
29	5	headspace-meditation-sleep	493145008
53	5	headspace-meditation-sleep	493145008
58	1	headspace-meditation-sleep	493145008
93	1	headspace-meditation-sleep	493145008
98	3	headspace-meditation-sleep	493145008
...
44571	4	relax-meditation-guided-mind	367506176
44575	3	relax-meditation-guided-mind	367506176
44603	1	relax-meditation-guided-mind	367506176
44618	1	relax-meditation-guided-mind	367506176
44624	3	relax-meditation-guided-mind	367506176

developerResponse

29

{'id': 14974073, 'body': 'We are sorry for the trouble with accessing your account! Could you please provide us with your email reference number so we can follow up with you?', 'modified': '2020-04-28T18:57:36Z'}

53

{'id': 16546248, 'body': 'So sorry to hear of any issues with the Sleepcasts! Please send our team a note at help@headspace.com so we can look into this for you.', 'modified': '2020-07-11T00:04:21Z'}

58

```

{'id': 17169865, 'body': "Thank you for your review regarding the Apple Watch
options, we'll be sure this feedback is heard by our team! If you have any
questions, feel free to get in touch with our team at help@headspace.com.",
'modified': '2020-08-09T20:35:33Z'}
93
{'id': 22831993, 'body': "We're sorry for the trouble with the Family Plan.
Since you purchased through Apple, you will have to wait until your subscription
expires, then you can purchase the Family Plan. Please send us a note to
help@headspace.com if you have any other questions. ", 'modified':
'2021-05-16T17:30:51Z'}
98
{'id': 18484659, 'body': "We're sorry about the trouble with the notifications.
Please send us a note to help@headspace.com and our team can further assist!",
'modified': '2020-10-14T19:56:00Z'}
...
...
44571 {'id': 11513811, 'body': "Hello, thanks for your review. Sorry to hear
about this, the app should restore your past purchases. Can you please tap the
Restore subscription button under Profile/Settings? This should solve the issue.
If it doesn't, contact us at support@ipnos.com and send us a screenshot from
your iTunes purchase history that shows which access has been purchased, so we
can sort this out.\u2028\n\nGabriel from Ipnos\n", 'modified':
'2019-10-28T14:19:47Z'}
44575
{'id': 13086567, 'body': 'Hello, thanks for your review! We try to offer as much
free content as possible but we also need to sell some features to keep on
developing the application! Yanik from Ipnos', 'modified':
'2020-01-23T18:49:47Z'}
44603
{'id': 13086742, 'body': 'Hello, thanks for your review! We try to offer as much
free content as possible but we also need to sell some features to keep on
developing the application! Yanik from Ipnos', 'modified':
'2020-01-23T19:03:44Z'}
44618
{'id': 13086654, 'body': 'Hello, thanks for your review. We are sorry to learn
that you are experiencing issues with our app, this is not expected behavior.
The problem you are describing is a known issue, it should be fixed in our next
update.\u2028 Yanik from Ipnos', 'modified': '2020-01-23T18:55:41Z'}
44624
{'id': 6642146, 'body': 'Hello J4m3$rawdogger, thanks for reaching out, we added
your suggestion to our community wishlist. Contact us at support@ipnos.com if
you have more suggestions for us!', 'modified': '2019-01-02T15:30:31Z'}

[4765 rows x 9 columns]

```

Upon first glance, we can see that the title, userName, date, review, rating and app_name could be relevant to this analysis. Since this analysis focuses on determining sentiments of the app users,

we will be primarily examining the title, review and rating columns.

4.1.1 Removing Null Values

```
[11]: # Check null values in review column
df[df['review'].isna()]
```

```
[11]:
      title \
17939 Trying to remove my review after request for refund was accepted

      isEdited  userName          date review  rating \
17939      True  Xwave500  2021-01-23 13:54:34    NaN      4

      app_name      app_id \
17939 talkspace-therapy-counseling  661829386

developerResponse
17939 {'id': 20485547, 'body': 'Hi there, we are so sorry to hear of your less
than optimal experience. Someone from our team would love to look into what
happened and do what we can to make this right by you. Please send us an email
at Feedback@Talkspace.com, we look forward to hearing from you. ', 'modified':
'2021-01-18T20:54:59Z'}
```

We have only a single review that is has a null value, and upon closer examination, we can see this is because the user was trying to remove the review. Since the title and rating columns alone are not a good indication of this user's sentiment, we will drop this single row.

```
[12]: # Remove row with null value for review
df = df[df['review'].isna()==False]
```

4.1.2 Addressing Duplicated Data

```
[13]: # Check for duplicated rows
df[df.duplicated(keep=False)].sort_values('title')
```

```
[13]:
      title  isEdited \
17116      5      False
17004      5      False
12417  Beginning      False
9182    Beginning      False
11782    Best app      False
12476    Best app      False
12136    Best app EVER!!!!!!!  False
12494    Best app EVER!!!!!!!  False
12246      Calm      False
12477      Calm      False
12478      Calm      False
```


11916	Calm	False
12106	Can not unlock	False
12442	Can not unlock	False
8565	Could do hardly anything	False
8568	Could do hardly anything	False
44570	Doing well	False
34690	Doing well	False
26098	Fabulous connects me to my soul!	False
26100	Fabulous connects me to my soul!	False
11465	Favorite App	False
12427	Favorite App	False
44212	Gonna delete	False
37217	Gonna delete	False
17092	Good karma	False
17137	Good karma	False
12392	H a r r y s t y l e a	False
11937	H a r r y s t y l e a	False
12447	I love it	False
12484	I love it	False
12157	I love it	False
11863	I love it	False
12421	In Spanish please the voice of her	False
9380	In Spanish please the voice of her	False
12298	Love	False
12454	Love	False
8569	No very good.	False
8566	No very good.	False
8570	Premium	False
8564	Premium	False
12479	Sleepy	False
11420	Sleepy	False
12474	Vital	False
12231	Vital	False
3935	WHY IS EVERYTHING LOCKED	False
808	WHY IS EVERYTHING LOCKED	False

	userName	date \
17116	kikiriggs	2021-03-30 12:04:53
17004	kikiriggs	2021-03-30 12:04:53
12417	Weather ghost	2021-01-05 00:29:35
9182	Weather ghost	2021-01-05 00:29:35
11782	Bam3322	2020-02-02 20:17:03
12476	Bam3322	2020-02-02 20:17:03
12136	idealighting	2020-04-02 00:39:55
12494	idealighting	2020-04-02 00:39:55
12246	Elyssa is a dog lover	2020-01-29 12:33:21
12477	Summer Dawsonson	2020-02-01 00:33:48

12478	Elyssa is a dog lover	2020-01-29 12:33:21
11916	Summer Dawson	2020-02-01 00:33:48
12106	Elenauionl	2020-07-22 21:04:15
12442	Elenauionl	2020-07-22 21:04:15
8565	pigletgirl2011	2020-06-20 05:50:08
8568	pigletgirl2011	2020-06-20 05:50:08
44570	Liz	2020-09-16 03:29:29
34690	Liz	2020-09-16 03:29:29
26098	justrefundme	2021-06-05 16:24:31
26100	justrefundme	2021-06-05 16:24:31
11465	ShelleBump	2020-10-30 03:18:04
12427	ShelleBump	2020-10-30 03:18:04
44212	hrhehehehehehehrheheh	2020-01-17 22:09:26
37217	hrhehehehehehehrheheh	2020-01-17 22:09:26
17092	raj2ray	2020-03-03 17:19:53
17137	raj2ray	2020-03-03 17:19:53
12392	avilene m	2020-07-10 15:18:59
11937	avilene m	2020-07-10 15:18:59
12447	MaryamEspahbodi	2020-07-04 06:04:46
12484	Zo	2020-01-15 00:11:15
12157	Zo	2020-01-15 00:11:15
11863	MaryamEspahbodi	2020-07-04 06:04:46
12421	nievedefebrero	2020-12-12 13:16:17
9380	nievedefebrero	2020-12-12 13:16:17
12298	bfxfbdfgfhmfgygj,hghgvsgdgi	2020-05-11 16:53:33
12454	bfxfbdfgfhmfgygj,hghgvsgdgi	2020-05-11 16:53:33
8569	Bookjhui	2020-06-10 00:10:12
8566	Bookjhui	2020-06-10 00:10:12
8570	LAK	2020-06-08 03:53:35
8564	LAK	2020-06-08 03:53:35
12479	hdbxifncjnsn hdjxn jdmccj nc	2020-01-26 05:36:07
11420	hdbxifncjnsn hdjxn jdmccj nc	2020-01-26 05:36:07
12474	Tgalluzzo	2020-02-11 12:12:15
12231	Tgalluzzo	2020-02-11 12:12:15
3935	Rock Lee Lee Rock	2020-12-15 05:26:40
808	Rock Lee Lee Rock	2020-12-15 05:26:40

review \

17116
 So lovely
 17004
 So lovely
 12417
 Just started but like it so far
 9182
 Just started but like it so far
 11782

Period
12476
Period
12136
Just awesome
12494
Just awesome
12246
I calm
12477
Great app I love it
12478
I calm
11916
Great app I love it
12106
I can not unlock Calm
12442
I can not unlock Calm
8565
I would like it if you could unlock more things on the app because I got this to help myself sleep better but I can't sleep any better.
8568
I would like it if you could unlock more things on the app because I got this to help myself sleep better but I can't sleep any better.
44570
So far this app is helping me and letting me journal in one single place and have it all organized.
34690
So far this app is helping me and letting me journal in one single place and have it all organized.
26098
I absolutely love Fabulous! It helps me relax, hear my own thoughts, introspect, and reminds me to practice gratitude and the importance of taking care of myself. It helps me connect to my soul.
26100
I absolutely love Fabulous! It helps me relax, hear my own thoughts, introspect, and reminds me to practice gratitude and the importance of taking care of myself. It helps me connect to my soul.
11465
Love Calm so much.
12427
Love Calm so much.
44212
I love it but I'm gonna delete it cause I need space
37217
I love it but I'm gonna delete it cause I need space

17092
Good karma
17137
Good karma
12392
b e d
11937
b e d
12447
Amazing
12484
I love Calm I LOVE it
12157
I love Calm I LOVE it
11863
Amazing
12421
What great if you put sonidos , música y cuenta cuentos en Español . Yo soy
Maestra , love reed and I counted historíes to my Spanish er kids in Uruguay.
Please I am very anxious in this days and need it .
9380
What great if you put sonidos , música y cuenta cuentos en Español . Yo soy
Maestra , love reed and I counted historíes to my Spanish er kids in Uruguay.
Please I am very anxious in this days and need it .
12298
Love it
12454
Love it
8569
In my opinion you can get most of this online for basically free. I should have
just tried it for a month instead of wasted the \$70 for the year. My fault, oh
well lesson learned. Wish I could get my money back. !
8566
In my opinion you can get most of this online for basically free. I should have
just tried it for a month instead of wasted the \$70 for the year. My fault, oh
well lesson learned. Wish I could get my money back. !
8570
I love the app, there just isn't much you can do without paying 60 dollars.
8564
I love the app, there just isn't much you can do without paying 60 dollars.
12479
It makes me very so sleepy
11420
It makes me very so sleepy
12474
Tamara is amazing
12231

Tamara is amazing

3935 I love this app, and it's overall great. I used to have little to no complaints, but now it's insane. I get that you need money for your business, but everything is locked. It's expensive and very sad. I used to listen to beachcomber every night until everything was locked and you needed to pay for it. It's fine to have some payment things, but at least leave 1/3 of them free. I don't want to pay a lot of money just to listen to music or a sleep cast.

808 I love this app, and it's overall great. I used to have little to no complaints, but now it's insane. I get that you need money for your business, but everything is locked. It's expensive and very sad. I used to listen to beachcomber every night until everything was locked and you needed to pay for it. It's fine to have some payment things, but at least leave 1/3 of them free. I don't want to pay a lot of money just to listen to music or a sleep cast.

	rating	app_name	app_id	developerResponse
17116	5	insight-timer-meditation-app	337472899	NaN
17004	5	insight-timer-meditation-app	337472899	NaN
12417	4	calm	571800810	NaN
9182	4	calm	571800810	NaN
11782	5	calm	571800810	NaN
12476	5	calm	571800810	NaN
12136	5	calm	571800810	NaN
12494	5	calm	571800810	NaN
12246	5	calm	571800810	NaN
12477	5	calm	571800810	NaN
12478	5	calm	571800810	NaN
11916	5	calm	571800810	NaN
12106	1	calm	571800810	NaN
12442	1	calm	571800810	NaN
8565	2	calm	571800810	NaN
8568	2	calm	571800810	NaN
44570	5	reflectly-journal-ai-diary	1241229134	NaN
34690	5	reflectly-journal-ai-diary	1241229134	NaN
26098	5	fabulous-daily-routine-planner	1203637303	NaN
26100	5	fabulous-daily-routine-planner	1203637303	NaN
11465	5	calm	571800810	NaN
12427	5	calm	571800810	NaN
44212	5	reflectly-journal-ai-diary	1241229134	NaN
37217	5	reflectly-journal-ai-diary	1241229134	NaN
17092	5	insight-timer-meditation-app	337472899	NaN
17137	5	insight-timer-meditation-app	337472899	NaN
12392	5	calm	571800810	NaN
11937	5	calm	571800810	NaN
12447	5	calm	571800810	NaN
12484	5	calm	571800810	NaN
12157	5	calm	571800810	NaN
11863	5	calm	571800810	NaN

12421	4		calm	571800810	NaN
9380	4		calm	571800810	NaN
12298	5		calm	571800810	NaN
12454	5		calm	571800810	NaN
8569	1		calm	571800810	NaN
8566	1		calm	571800810	NaN
8570	4		calm	571800810	NaN
8564	4		calm	571800810	NaN
12479	5		calm	571800810	NaN
11420	5		calm	571800810	NaN
12474	5		calm	571800810	NaN
12231	5		calm	571800810	NaN
3935	2	headspace-meditation-sleep		493145008	NaN
808	2	headspace-meditation-sleep		493145008	NaN

Although it's not clear why there are perfectly duplicated observations, we will drop duplicated instances of these observations to prevent our data from being biased.

```
[14]: #Remove duplicated rows
df = df.drop_duplicates()
```

Although we have dropped perfectly duplicated rows, we also want to look into whether there were users who have written multiple reviews for the same app.

```
[15]: # Check for multiple reviews for same app from same user
df[df.duplicated(['userName', 'app_name'], keep=False)]
```

```
[15]:
```

	title	isEdited	userName \
18882	Worth it, but with a flaw for me	False	Hobbit of the Shire
18911	Love app, but lost data!	False	Hobbit of the Shire

	date \
18882	2021-01-24 18:03:56
18911	2021-01-19 21:15:26

	review \
18882	I found MindDoc (previously Moodpath) when I was struggling with my mental health and wanted to track my emotions and journal my thoughts throughout the day. This app is perfect for that. \n\nCurrently, the free version asks you a few questions 3x a day, allows you to record your emotional state and any thoughts, and gives you a report every two weeks. \n\nWhat appeals to me is the streamlined nature of the app. It has additional content for those interested, but for me, the free version fufills all that I need it to.\n\nThe only flaw is that I had data for about 6 months before they changed the name and structure of the app, resulting in the loss of all my journal enteries. Unfortunately, they were unable to retrieve the data due to the way the app was structured before. Other than this flaw, I really enjoy this app and would recommend it to others.
18911	

I've really enjoyed using this app. It's very streamlined, just a mood tracker and journal. Perfect!\n\nUnfortunately, when Moodpath switched to MindDoc, I got logged out (forgot the password) and lost my data. It won't send me an email to reset it. I'm seeing other reviewers experiencing the same problem. I've already contacted the support email with my issue and still waiting on a response.\n\nI'll update this review to five stars when it gets fixed because it's worked great up till this point. Thank you!

	rating	app_name	app_id	\
18882	4	minddoc-your-companion	1052216403	
18911	2	minddoc-your-companion	1052216403	

developerResponse

18882

NaN

18911 {'id': 20551815, 'body': 'Hello, thank you for reaching out. We are very sorry you lost your data! We will contact you via mail as soon as we manage to bring it back. Our team is working hard to solve this inconvenient problem. Best wishes and thank you for your patience, your MindDoc team', 'modified': '2021-01-21T11:43:06Z'}

We only have one user who left multiple reviews, and both reviews seem to be valid observations of how the user felt about the app at different times.

4.1.3 Feature Engineering

Because both the title and review columns hold text that are relevant to our analysis, we will combine them into one column. This will ensure that we are able to simultaneously examine information in the title and review columns.

```
[16]: # Combine title and review column to get full review text
df['combined_text'] = df['title'] + ". " + df['review']
df
```

```
[16]:
```

	title	isEdited	\
0	This is the BEST app by far for mindfulness	False	
1	Facing down the rabbit hole	False	
2	A Bright Spot in a Dark World	False	
3	The most useful thing I've done in 2020	False	
4	It's a Miracle App!	False	
...	
44717	Me encanta lo mejor para meditar	False	
44718	Excelente	False	
44719	Love this app	False	
44720	Disappointed	False	
44721	FALSE ADVERTISEMENT subscription terms	False	

userName	date	\
----------	------	---

0	Crazy Yorkie Lady	2021-02-22 18:13:54
1	KLC-MHFL	2021-02-12 12:42:11
2	Ashabashley	2021-02-11 06:22:38
3	Kindly38	2021-01-18 17:14:21
4	jlhuggins	2021-02-08 19:19:10
...
44717	JLucret	2020-04-06 23:34:12
44718	mata rato	2020-03-28 06:01:54
44719	Fierce Girl	2020-03-24 13:44:13
44720	Irkesome	2020-03-07 17:29:09
44721	ksavv916	2020-02-18 02:08:56

review \

0 It's almost embarrassing to say this, but I have been a "meditator" for about 30 years and I have to say that I never really got very much out of doing it. I tried so hard, but I'm finding out I never truly understood that "trying" to be a good meditator or mindful person, was missing what doing this was about. What I'm learning through this app is that "mindfulness" is about me - understanding myself first, and it's a lot about acceptance. So with that understanding and acceptance I can begin to see "things" so much more clearly and why previous ways of thinking, particularly about myself, were holding me back from living my life and loving myself. And I'm finding that the more I practice self acceptance, the fears about living fully and going after what I want out of my life, are disappearing. \n\nThe previous version of myself is rapidly becoming someone I only used to know, and this new ME is much more in line with who I'd always wanted to be. I'm able to try and do so much more, all while in the comfort that I am lovable, even if I fail at something. I simply dust myself off and try some more.\n\nThis app has completely changed me and the way I'm able to appreciate all the good that's in my life. And I'm worthy of going after the things I want for my life, while freeing me up to share my gifts with others and society in general.\n\nIf this is where you find your life stalling out, this app will bring you back to yourself.

1 I discovered Headspace on Netflix. My life is a series of invitations to succumb to fear. My husband is a practicing psychologist who is on peritoneal dialysis for 10 hours every night and works remotely with patients for about 10 hours every day without complaint. We have been waiting for a kidney for two years. I am optimistic by nature but even if this was the only focus for concern it would be a lot. Unhelpfully, the list goes on with equivalent issues. I can't work; I need to be on call because I am the go-to for support. I occupy my mind and time writing a book and renovating every room of our home one at a time amidst being available for nine blended (mostly grown) children and 5 grandkids, many of whom have special needs. I am busy but isolated like so many due to covid-19. Trying the Headspace series felt like meeting air; filling the natural and effortless need to breathe. It has provided me with a new lens through which to focus on challenges and see potential. I incorporate it into my waking routine at dawn and it positions me for whatever happens throughout the day, and differentiates a space that is my own. I turn 61

this month and for maybe the first time in a while, life feels lighter and brighter. It's not about what you fear or can't control or even the fact of fear. It's about the lens through which you interpret your fears or challenges and making them your strengths and allies. Thank you Andy :-)

2 You won't regret buying a Headspace subscription. I often joke with people that Headspace is the one thing I would want to take with me to a desert island. If I were to cancel all of my subscriptions today, Headspace is the one exception I would make. I'm a high school English teacher, and our administration put together a month-long mindfulness education program for our students shortly after I subscribed. I had already raved to many of them about the app and the benefits of taking a couple of minutes out of their day to sit and be present. Lo and behold, Andy pops up on the next slide as our Ted Talk speaker of the day! On a more serious note, I suffer from debilitating chronic pain. I'm currently on long-term disability and a leave of absence from the classroom. Mindfulness, and the courses on managing pain, have been a tremendous help in my journey. Not only that, but I struggle with insomnia. My favorite part of Headspace, because of my (admittedly) vivid imagination, are the sleepcasts. I love falling asleep to Rainday Antiques. I can't imagine doing anything but that ever again. I have learned to manage stress and anxiety in a way that I had never thought was possible. I have meditated in the waiting room of The Cleveland Clinic before my appointments. I once laughed at the idea of mindfulness and meditation. Now it's an integral part of my life. It's part of my journey, and I think everyone could benefit from it.

3

This app has given me tools to deal with constant change. This past year, I've needed to pivot every day to face some new challenge outside as an essential worker, and at home in isolation with my partner. The tools I learned to use to be present in the moment and to accept change has benefited my partner and colleagues at work as much as my own peace of mind. I have very little control over my environment, but I can see my reaction, accept the emotions that rise up, and move on instead of getting lost or bound by them. \nThis app has also benefited me by helping me get through insomnia. Sometimes it helps me get to sleep in the first place. Mostly it gives me tools to use to calm my mind and body when I wake up in the night buffeted by anxiety or lingering frustration from the day. \nLastly, this app has given me a language to express the emotional upheaval and troubling questions that I used to just call anxiety. Naming what makes me feel upset has brought clarity to what can feel like a tempest at the time and allow me to look at my thoughts or, rather, face them. After acknowledging them, I can let them go. Sometimes I'm better at that than at other times. \nRegular exercise strengthens my focus, my ability to pivot and refocus, and builds roads of calm and quiet that the brain finds easier to fall into and travel due to constant use.

4

This app has helped me so much! When I am suffering from stress, health anxiety, or fibro/stress related body pains I turn to this app and go through some meditations with the teachers. It's an excellent way for me to relax. Calm my mind. Stop thinking about myself. And to recover in the moment. Hopefully in

time this will help me heal, as previously I've just been getting worse and worse without any therapy since my therapist's practice permanently shutdown from COVID. This has been a great help and is something I can just pull out of my pocket and sit in my bedroom for a bit to calm down. Well worth the 1 year subscription price I've paid. Thank you! \n\n*There is only one thing that I'd change. This would be for some of the workout audios. I need to walk a lot for a heart condition I have (which is a big cause of my health anxiety), and I'd love to see the audios for walks and runs to be a bit longer. The longest run audio I see is 20 minutes, when recommended cardio workout lengths realistically should be at least 30 minutes. I'd really like to have an audio similar to the bedtime audios that are lengthy, have soothing talk, and tips to reduce anxiety made to be listened while walking. And if there could be a variety to choose from. Overall this would be the only change I can think of. An hour long walk audio to help calm my mind.

...

...

44717

Es la mejor aplicación se las recomiendo.

44718

Lo recomiendo

44719

Great app.-would love it if the sounds were organized in a way that makes sense.

44720

You get exactly two meditations before you pay 60.00 per year.

44721

As stated in the App Information section WORD FOR WORD: "RelaxMeditation also offers a lifetime subscription for \$19.99 which is paid for by a one-off upfront payment with unlimited access to RelaxMeditation forever." \n\nMy message to App Support:\n"I'm trying to purchase the \$19.99 lifetime / unlimited subscription but can't find where / how to?" \n\nTheir response: \n"Hello Kristen, thank you for reaching out to us! I really apologize about this. The offers and prices displayed in the Store are not accurate, as they display all of our past and current offers for all of our apps. \n[...] We do have a special right now, and it offers 60% off the regular price, for 99.99 USD. [...] We also offer a yearly subscription, for 59.99\$, and it includes a 7-day free trial period.[..]"

\n\n60% off- at \$99. \nNo thanks. Good luck with that.

	rating	app_name	app_id	developerResponse	\
0	5	headspace-meditation-sleep	493145008	NaN	
1	5	headspace-meditation-sleep	493145008	NaN	
2	5	headspace-meditation-sleep	493145008	NaN	
3	5	headspace-meditation-sleep	493145008	NaN	
4	5	headspace-meditation-sleep	493145008	NaN	
...	
44717	5	relax-meditation-guided-mind	367506176	NaN	
44718	5	relax-meditation-guided-mind	367506176	NaN	
44719	4	relax-meditation-guided-mind	367506176	NaN	

44720	2	relax-meditation-guided-mind	367506176	NaN
44721	1	relax-meditation-guided-mind	367506176	NaN

combined_text

0 This is the BEST app by far for mindfulness. It's almost embarrassing to say this, but I have been a "meditator" for about 30 years and I have to say that I never really got very much out of doing it. I tried so hard, but I'm finding out I never truly understood that "trying" to be a good meditator or mindful person, was missing what doing this was about. What I'm learning through this app is that "mindfulness" is about me - understanding myself first, and it's a lot about acceptance. So with that understanding and acceptance I can begin to see "things" so much more clearly and why previous ways of thinking, particularly about myself, were holding me back from living my life and loving myself. And I'm finding that the more I practice self acceptance, the fears about living fully and going after what I want out of my life, are disappearing. \n\nThe previous version of myself is rapidly becoming someone I only used to know, and this new ME is much more in line with who I'd always wanted to be. I'm able to try and do so much more, all while in the comfort that I am lovable, even if I fail at something. I simply dust myself off and try some more.\n\nThis app has completely changed me and the way I'm able to appreciate all the good that's in my life. And I'm worthy of going after the things I want for my life, while freeing me up to share my gifts with others and society in general.\n\nIf this is where you find your life stalling out, this app will bring you back to yourself.

1 Facing down the rabbit hole. I discovered Headspace on Netflix. My life is a series of invitations to succumb to fear. My husband is a practicing psychologist who is on peritoneal dialysis for 10 hours every night and works remotely with patients for about 10 hours every day without complaint. We have been waiting for a kidney for two years. I am optimistic by nature but even if this was the only focus for concern it would be a lot. Unhelpfully, the list goes on with equivalent issues. I can't work; I need to be on call because I am the go-to for support. I occupy my mind and time writing a book and renovating every room of our home one at a time amidst being available for nine blended (mostly grown) children and 5 grandkids, many of whom have special needs. I am busy but isolated like so many due to covid-19. Trying the Headspace series felt like meeting air; filling the natural and effortless need to breathe. It has provided me with a new lens through which to focus on challenges and see potential. I incorporate it into my waking routine at dawn and it positions me for whatever happens throughout the day, and differentiates a space that is my own. I turn 61 this month and for maybe the first time in a while, life feels lighter and brighter. It's not about what you fear or can't control or even the fact of fear. It's about the lens through which you interpret your fears or challenges and making them your strengths and allies. Thank you Andy :-)

2 A Bright Spot in a Dark World. You won't regret buying a Headspace subscription. I often joke with people that Headspace is the one thing I would want to take with me to a desert island. If I were to cancel

all of my subscriptions today, Headspace is the one exception I would make. I'm a high school English teacher, and our administration put together a month-long mindfulness education program for our students shortly after I subscribed. I had already raved to many of them about the app and the benefits of taking a couple of minutes out of their day to sit and be present. Lo and behold, Andy pops up on the next slide as our Ted Talk speaker of the day! On a more serious note, I suffer from debilitating chronic pain. I'm currently on long-term disability and a leave of absence from the classroom. Mindfulness, and the courses on managing pain, have been a tremendous help in my journey. Not only that, but I struggle with insomnia. My favorite part of Headspace, because of my (admittedly) vivid imagination, are the sleepcasts. I love falling asleep to Rainday Antiques. I can't imagine doing anything but that ever again. I have learned to manage stress and anxiety in a way that I had never thought was possible. I have meditated in the waiting room of The Cleveland Clinic before my appointments. I once laughed at the idea of mindfulness and meditation. Now it's an integral part of my life. It's part of my journey, and I think everyone could benefit from it.

3

The most useful thing I've done in 2020. This app has given me tools to deal with constant change. This past year, I've needed to pivot every day to face some new challenge outside as an essential worker, and at home in isolation with my partner. The tools I learned to use to be present in the moment and to accept change has benefited my partner and colleagues at work as much as my own peace of mind. I have very little control over my environment, but I can see my reaction, accept the emotions that rise up, and move on instead of getting lost or bound by them. \nThis app has also benefited me by helping me get through insomnia. Sometimes it helps me get to sleep in the first place. Mostly it gives me tools to use to calm my mind and body when I wake up in the night buffeted by anxiety or lingering frustration from the day. \nLastly, this app has given me a language to express the emotional upheaval and troubling questions that I used to just call anxiety. Naming what makes me feel upset has brought clarity to what can feel like a tempest at the time and allow me to look at my thoughts or, rather, face them. After acknowledging them, I can let them go. Sometimes I'm better at that than at other times. \nRegular exercise strengthens my focus, my ability to pivot and refocus, and builds roads of calm and quiet that the brain finds easier to fall into and travel due to constant use.

4

It's a Miracle App!. This app has helped me so much! When I am suffering from stress, health anxiety, or fibro/stress related body pains I turn to this app and go through some meditations with the teachers. It's an excellent way for me to relax. Calm my mind. Stop thinking about myself. And to recover in the moment. Hopefully in time this will help me heal, as previously I've just been getting worse and worse without any therapy since my therapist's practice permanently shutdown from COVID. This has been a great help and is something I can just pull out of my pocket and sit in my bedroom for a bit to calm down. Well worth the 1 year subscription price I've paid. Thank you! \n\n*There is only one thing that I'd change. This would be for some of the workout audios. I

need to walk a lot for a heart condition I have (which is a big cause of my health anxiety), and I'd love to see the audios for walks and runs to be a bit longer. The longest run audio I see is 20 minutes, when recommended cardio workout lengths realistically should be at least 30 minutes. I'd really like to have an audio similar to the bedtime audios that are lengthy, have soothing talk, and tips to reduce anxiety made to be listened while walking. And if there could be a variety to choose from. Overall this would be the only change I can think of. An hour long walk audio to help calm my mind.

...

...

44717

Me encanta lo mejor para meditar. Es la mejor aplicación se las recomiendo.

44718

Excelente. Lo recomiendo

44719

Love this app. Great app.-would love it if the sounds were organized in a way that makes sense.

44720

Disappointed. You get exactly two meditations before you pay 60.00 per year.

44721

FALSE ADVERTISEMENT subscription terms. As stated in the App Information section WORD FOR WORD: "RelaxMeditation also offers a lifetime subscription for \$19.99 which is paid for by a one-off upfront payment with unlimited access to RelaxMeditation forever." \n\nMy message to App Support: \n"I'm trying to purchase the \$19.99 lifetime / unlimited subscription but can't find where / how to?" \n\nTheir response: \n"Hello Kristen, thank you for reaching out to us! I really apologize about this. The offers and prices displayed in the Store are not accurate, as they display all of our past and current offers for all of our apps. \n[...] We do have a special right now, and it offers 60% off the regular price, for 99.99 USD. [...] We also offer a yearly subscription, for 59.99\$, and it includes a 7-day free trial period.[...]" \n\n60% off- at \$99. \nNo thanks. Good luck with that.

[44698 rows x 10 columns]

4.1.4 Additional Text Formatting

Before moving onto exploring our data, we will address the issue of new-line characters that we can find in some of the reviews.

```
[17]: # Create a variable "corpus" containing all text
      corpus = df['combined_text'].to_list()
```

```
[18]: # Examine review with new-line character
      corpus[1000]
```

```
[18]: 'It made me anxious. This is how anxious I am that when I noticed the "managing anxiety" course was free, but expired soon, I tried to cram as much of the 30-day session in a couple days. I am that broke, thanks to covid, that the thought of not having time to do the course before they lock it up again, stressed me out. I had to uninstall it. \n\nIt's great, just not for me at the moment because I cannot afford it.'
```

```
[19]: # Replace all new-line characters with space
df['combined_text'] = df['combined_text'].replace(r'\n', ' ', regex=True)
# Create a variable "corpus" containing all text and verify
corpus = df['combined_text'].to_list()
corpus[1000]
```

```
[19]: 'It made me anxious. This is how anxious I am that when I noticed the "managing anxiety" course was free, but expired soon, I tried to cram as much of the 30-day session in a couple days. I am that broke, thanks to covid, that the thought of not having time to do the course before they lock it up again, stressed me out. I had to uninstall it. It's great, just not for me at the moment because I cannot afford it.'
```

Now that we've formatted our text, we can move on to the Exploratory Data Analysis.

5 EXPLORE

In this section, we will examine more characteristics of our data including the distribution of ratings and sentiments, as well as any insights we can extract via word frequencies. In order to accurately analyze word frequencies, we want to make sure that we are appropriately preprocessing each word via lemmatization and tokenization. We will then end this section by visualizing our positive, neutral and negative corpora in the form of wordclouds.

5.0.1 Data Distribution

We begin the exploration phase by examining the distribution of ratings and review counts across the apps.

```
[20]: # Check distribution of reviews for each app
df['app_name'].value_counts(normalize=True)
```

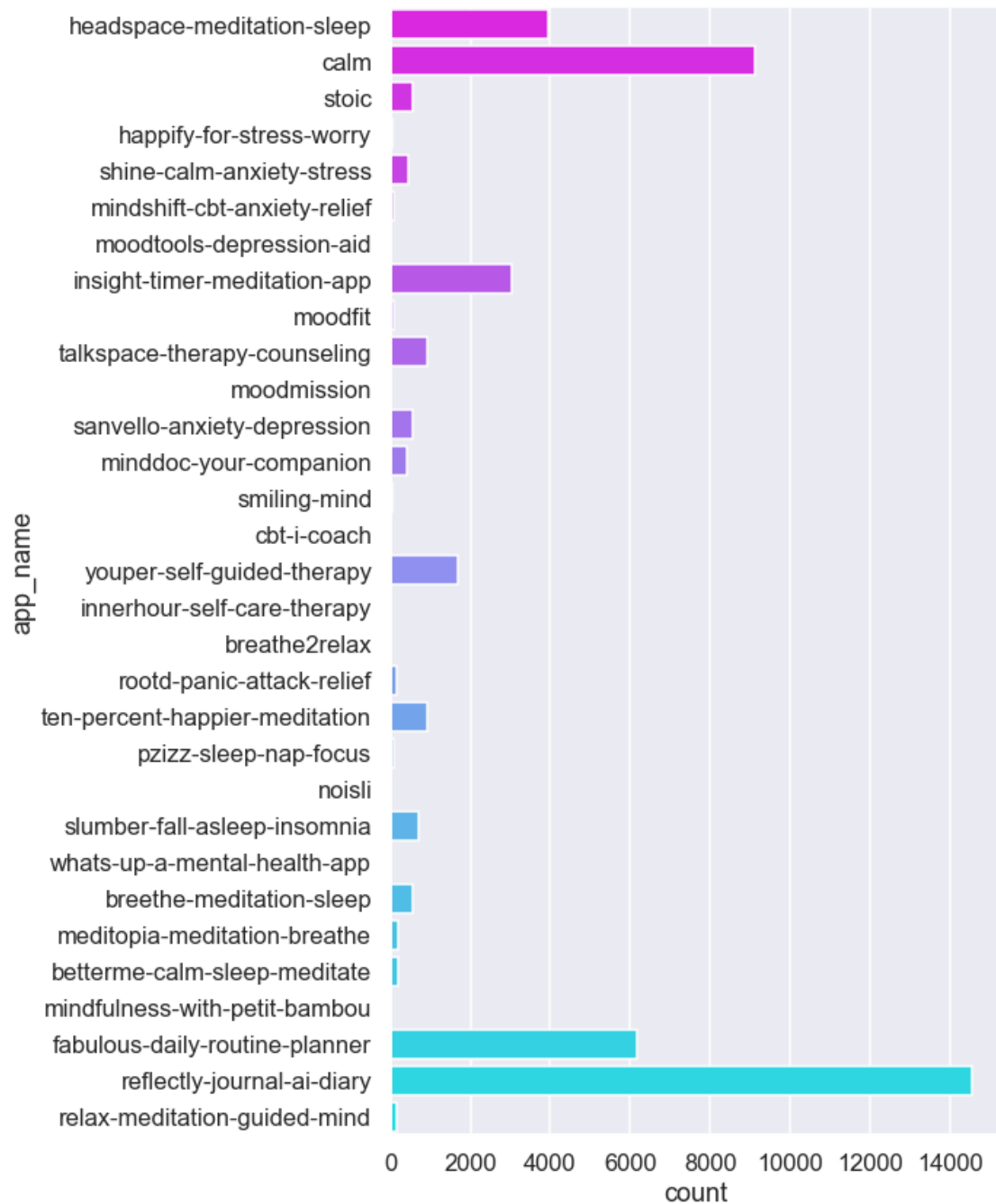
```
[20]: reflectly-journal-ai-diary      0.325809
      calm                        0.204193
      fabulous-daily-routine-planner 0.138015
      headspace-meditation-sleep    0.088035
      insight-timer-meditation-app  0.067609
      youper-self-guided-therapy    0.037407
      talkspace-therapy-counseling  0.020694
      ten-percent-happier-meditation 0.020381
      slumber-fall-asleep-insomnia  0.015616
      sanvello-anxiety-depression    0.012797
```

stoic	0.012350
breathe-meditation-sleep	0.012215
shine-calm-anxiety-stress	0.010135
minddoc-your-companion	0.009464
betterme-calm-sleep-meditate	0.004474
meditopia-meditation-breathe	0.003982
relax-meditation-guided-mind	0.003378
rootd-panic-attack-relief	0.003311
moodfit	0.001924
pzizz-sleep-nap-focus	0.001790
mindshift-cbt-anxiety-relief	0.001611
smiling-mind	0.001454
happify-for-stress-worry	0.000828
mindfulness-with-petit-bambou	0.000604
whats-up-a-mental-health-app	0.000515
noisli	0.000447
breathe2relax	0.000425
cbt-i-coach	0.000291
moodtools-depression-aid	0.000157
moodmission	0.000045
innerhour-self-care-therapy	0.000045

Name: app_name, dtype: float64

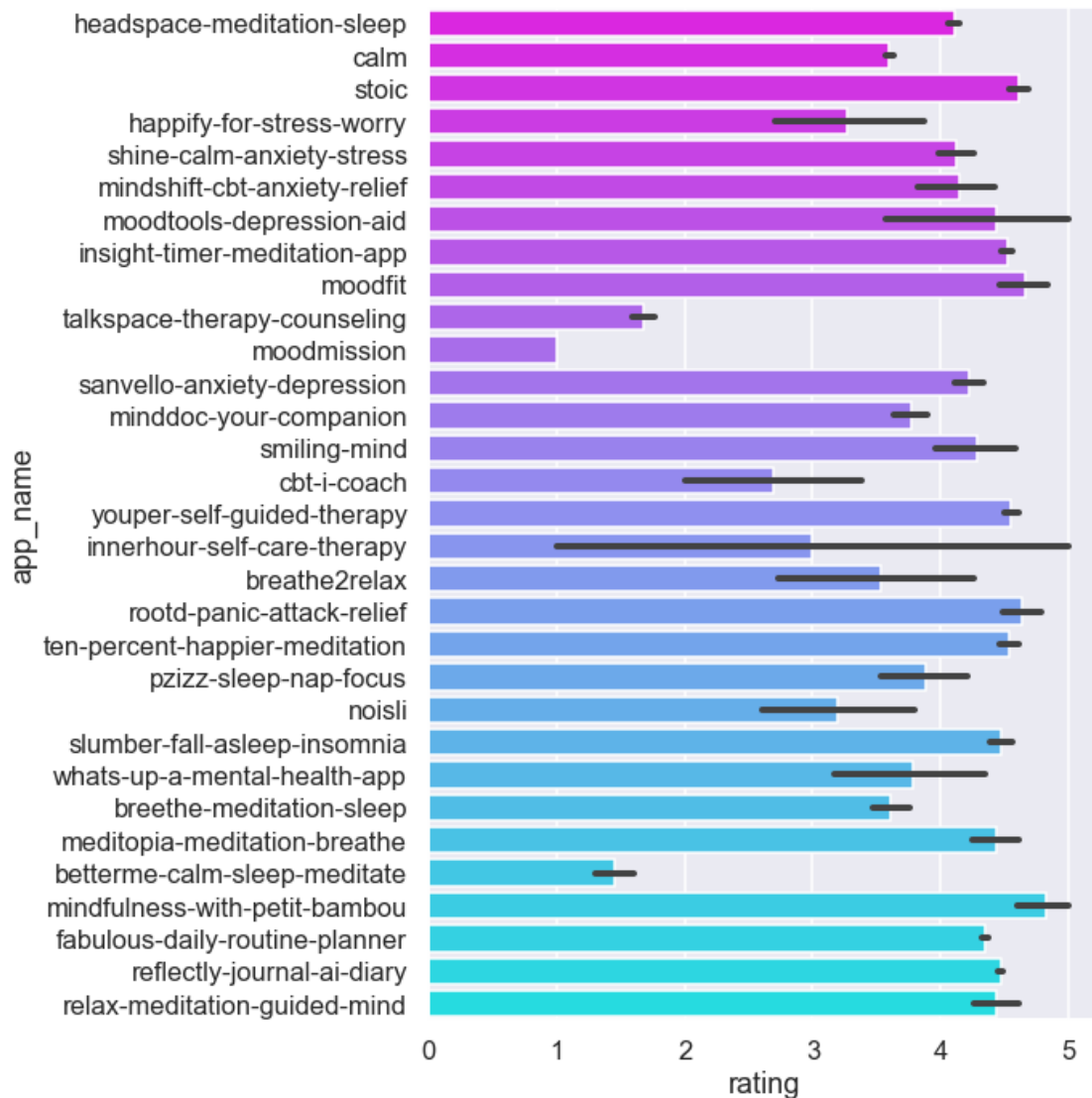
```
[21]: # Plot distribution of review count by app name
fig, ax = plt.subplots(figsize=(10,12))

sns.countplot(data=df, y='app_name', palette='cool_r', ax=ax, orient='h')
plt.tight_layout()
```



```
[22]: # Plot average rating by app name
fig, ax = plt.subplots(figsize=(10,10))

sns.barplot(data=df, x='rating', y='app_name', palette='cool_r',
            ax=ax, orient='h')
plt.tight_layout()
```

```
[23]: # Check distribution of ratings across all apps
df['rating'].value_counts()
```

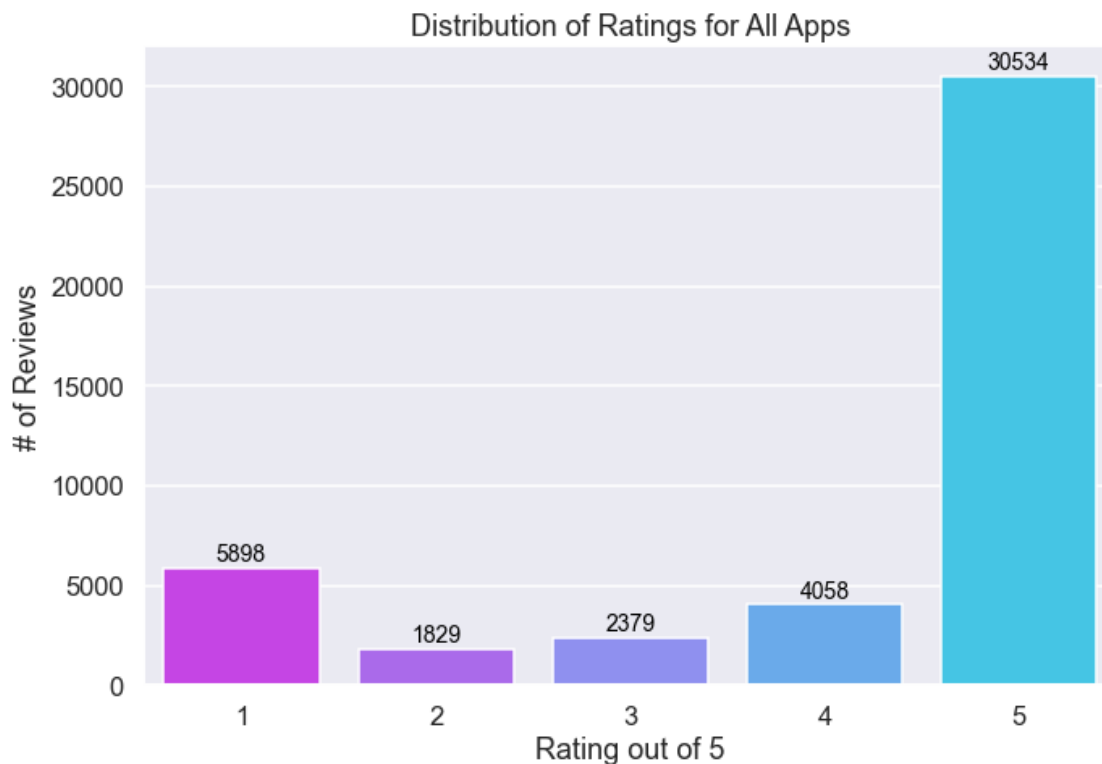
```
[23]: 5    30534
      1     5898
      4     4058
      3     2379
      2     1829
      Name: rating, dtype: int64
```

```
[24]: # Check distribution of ratings across all apps in percentages
df['rating'].value_counts(normalize=True)
```

```
[24]: 5    0.683118
      1    0.131952
      4    0.090787
      3    0.053224
      2    0.040919
      Name: rating, dtype: float64
```

```
[25]: # Plot distribution of ratings across all apps
fig, ax = plt.subplots(figsize=(10,7))
sns.countplot(x='rating', data=df, ax=ax, palette='cool_r')
ax.set_title('Distribution of Ratings for All Apps')
ax.set_ylabel('# of Reviews')
ax.set_xlabel('Rating out of 5')

for p in ax.patches:
    ax.annotate("%.0f" % p.get_height(), (p.get_x() + p.get_width() / 2.,
                                          p.get_height()),ha='center',
                va='center', fontsize=14, color='black', xytext=(0, 8),
                textcoords='offset points')
fig.tight_layout();
```



```
[26]: # Create dataframe with data for top five apps in terms of review count
distribution_df = df[df['app_name'].isin(['headspace-meditation-sleep', 'calm',
                                         'insight-timer-meditation-app',
                                         'fabulous-daily-routine-planner',
                                         'reflectly-journal-ai-diary'])]

# Calculate average mean rating across top five apps
distribution_df = distribution_df.groupby('app_name')['rating']\
    .value_counts(normalize=True).to_frame()

# Rename percentage column and app_name column
distribution_df.columns = ['percentage']

# Reset index, rename app column and display result
distribution_df.reset_index(inplace=True)
distribution_df.columns = ['App', 'rating', 'percentage']
distribution_df
```

```
[26]:
```

	App	rating	percentage
0	calm	5	0.527994
1	calm	1	0.233154
2	calm	4	0.084475
3	calm	3	0.079873
4	calm	2	0.074504
5	fabulous-daily-routine-planner	5	0.717134
6	fabulous-daily-routine-planner	4	0.124980
7	fabulous-daily-routine-planner	1	0.092884
8	fabulous-daily-routine-planner	3	0.036473
9	fabulous-daily-routine-planner	2	0.028530
10	headspace-meditation-sleep	5	0.670394
11	headspace-meditation-sleep	1	0.139771
12	headspace-meditation-sleep	4	0.084371
13	headspace-meditation-sleep	3	0.063532
14	headspace-meditation-sleep	2	0.041931
15	insight-timer-meditation-app	5	0.818994
16	insight-timer-meditation-app	1	0.069821
17	insight-timer-meditation-app	4	0.044672
18	insight-timer-meditation-app	3	0.038385
19	insight-timer-meditation-app	2	0.028127
20	reflectly-journal-ai-diary	5	0.761450
21	reflectly-journal-ai-diary	4	0.100941
22	reflectly-journal-ai-diary	1	0.064822
23	reflectly-journal-ai-diary	3	0.047037
24	reflectly-journal-ai-diary	2	0.025750

```
[27]: # Create map dictionary to format app names
app_name_map = {'calm': 'Calm',
```

```

        'fabulous-daily-routine-planner': 'Fabulous: Daily Routine_
↪Planner',
        'headspace-meditation-sleep': 'Headspace',
        'insight-timer-meditation-app': 'Insight Timer',
        'reflectly-journal-ai-diary': "Reflectly Journal & AI Diary"}

```

```

[28]: # Map unformatted app names to formatted app names and display
distribution_df['App'] = distribution_df['App'].map(app_name_map)
distribution_df

```

```

[28]:

```

	App	rating	percentage
0	Calm	5	0.527994
1	Calm	1	0.233154
2	Calm	4	0.084475
3	Calm	3	0.079873
4	Calm	2	0.074504
5	Fabulous: Daily Routine Planner	5	0.717134
6	Fabulous: Daily Routine Planner	4	0.124980
7	Fabulous: Daily Routine Planner	1	0.092884
8	Fabulous: Daily Routine Planner	3	0.036473
9	Fabulous: Daily Routine Planner	2	0.028530
10	Headspace	5	0.670394
11	Headspace	1	0.139771
12	Headspace	4	0.084371
13	Headspace	3	0.063532
14	Headspace	2	0.041931
15	Insight Timer	5	0.818994
16	Insight Timer	1	0.069821
17	Insight Timer	4	0.044672
18	Insight Timer	3	0.038385
19	Insight Timer	2	0.028127
20	Reflectly Journal & AI Diary	5	0.761450
21	Reflectly Journal & AI Diary	4	0.100941
22	Reflectly Journal & AI Diary	1	0.064822
23	Reflectly Journal & AI Diary	3	0.047037
24	Reflectly Journal & AI Diary	2	0.025750

```

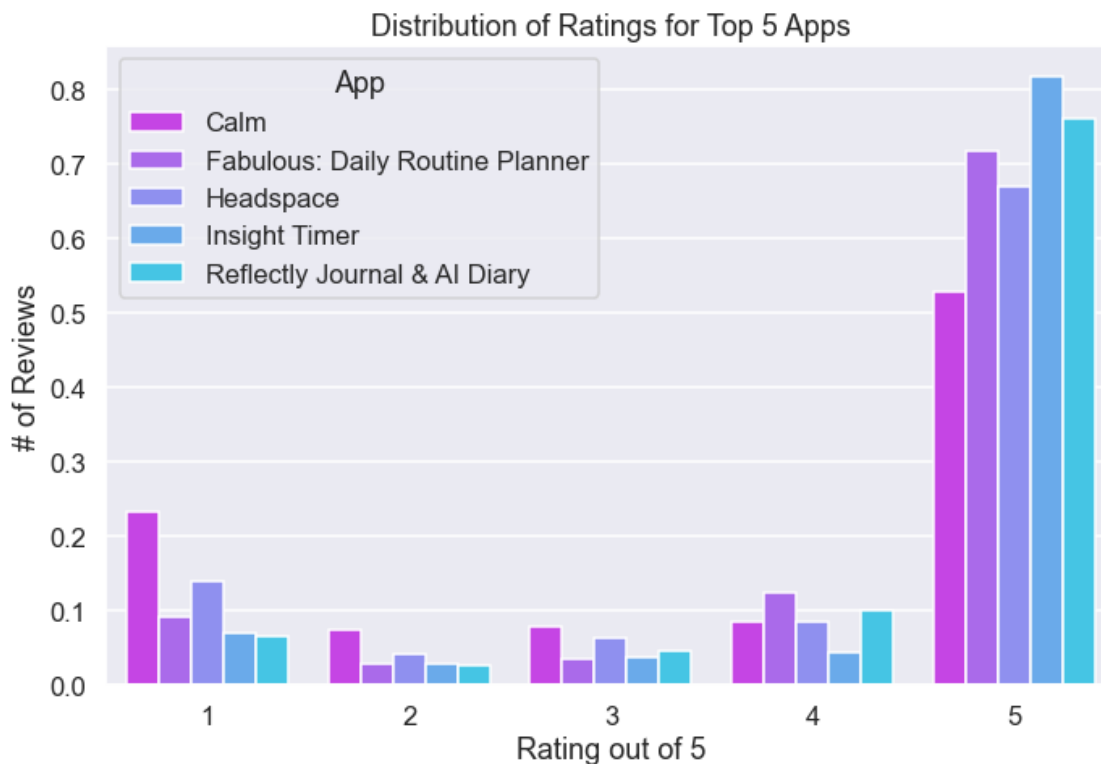
[29]: # Initialize figure
fig, ax = plt.subplots(figsize=(10,7))

# Create Barplot
sns.barplot(x='rating', y='percentage', hue='App', data=distribution_df,
            ax=ax, palette='cool_r')

# Format barplot
ax.set_title('Distribution of Ratings for Top 5 Apps')
ax.set_ylabel('# of Reviews')

```

```
ax.set_xlabel('Rating out of 5')
fig.tight_layout();
```



We can gather from the above plots that users are much more likely to leave positive 5-star ratings than they are to leave lower ratings and that the number of 1-star ratings is higher than ratings from 2-4 stars. We can also see that although there are some differences in distribution between apps, they generally follow a similar distribution of ratings.

5.0.2 Target Engineering

Here, we will create our target variables based on the ratings given by app users. First, we begin with a multi-class sentiment target and then move onto a version of a binary sentiment target where 4 stars and above are considered to be positive, as well as an alternative version where 3 stars and above are positive.

Multi-Class Sentiments

```
[30]: # Make new multiclass sentiment column based on rating
# 4 and above are positive, 3 is neutral, 1-2 are negative
conditions = [df['rating'] >= 4,
              df['rating'] == 3,
              df['rating'] <= 2,]
values = [2, 1, 0]
```

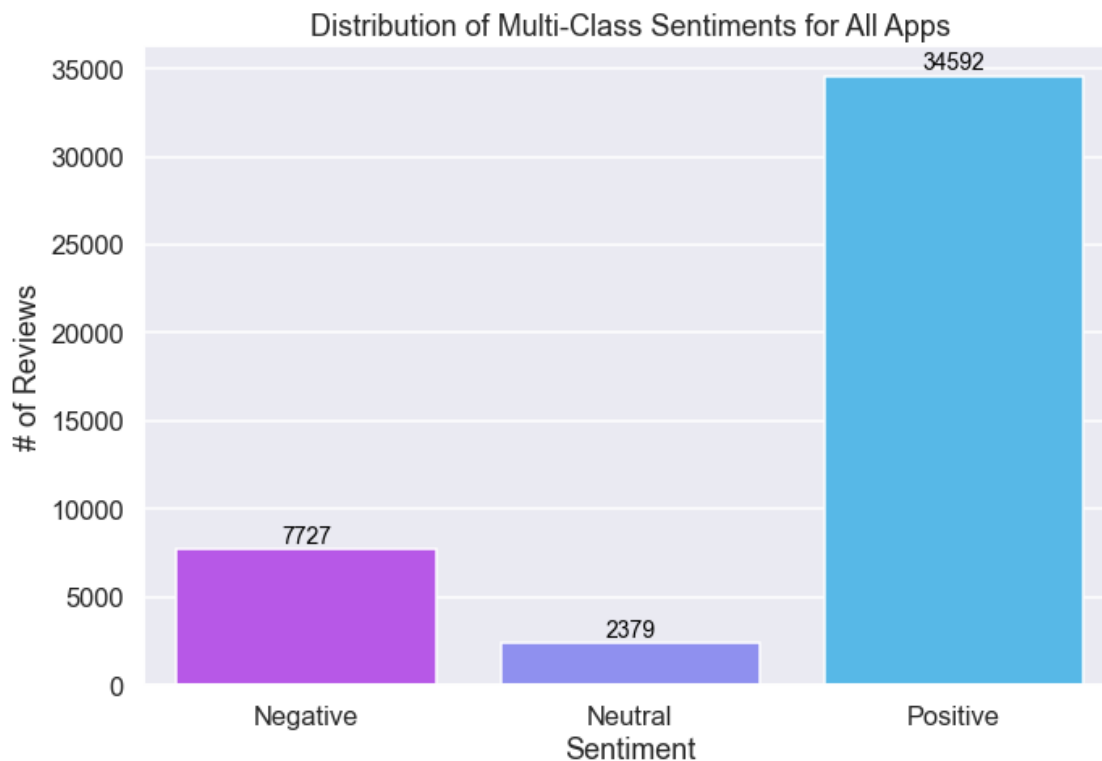
```
df['multi_sent'] = np.select(conditions, values)
```

```
[31]: # Plot distribution of multiclass sentiments across all apps
fig, ax = plt.subplots(figsize=(10,7))

sns.countplot(x='multi_sent', data=df, ax=ax, palette='cool_r')

ax.set_title('Distribution of Multi-Class Sentiments for All Apps')
ax.set_ylabel('# of Reviews')
ax.set_xlabel('Sentiment')
ax.set_xticklabels(['Negative', 'Neutral', 'Positive'])

for p in ax.patches:
    ax.annotate("%.0f" % p.get_height(), (p.get_x() + p.get_width() / 2.,
                                          p.get_height()), ha='center',
              va='center', fontsize=14, color='black', xytext=(0, 8),
              textcoords='offset points')
fig.tight_layout();
```



Binary Sentiments

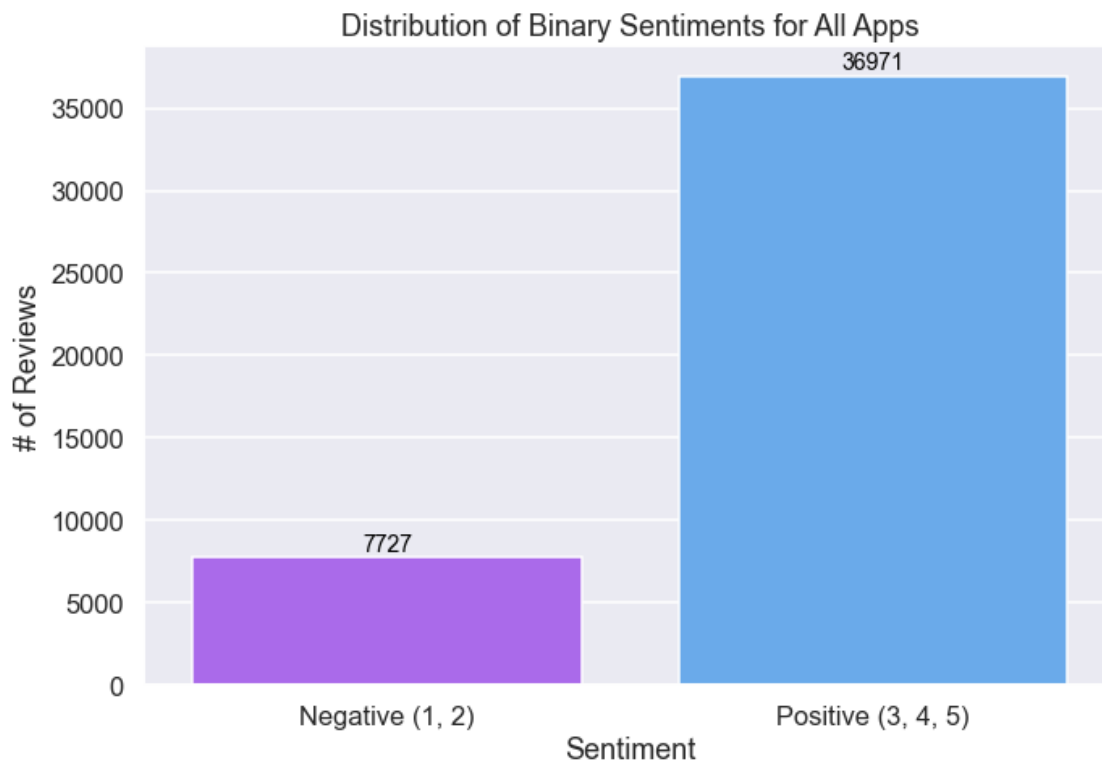
```
[32]: # Make new binary sentiment column based on ratings
# 3 and above are positive, below 3 is negative
conditions = [df['rating'] >= 3,
              df['rating'] < 3]

values = [1,0]

df['sent_bin_3up'] = np.select(conditions, values)

[33]: # Plot distribution of binary sentiments across all apps
fig, ax = plt.subplots(figsize=(10,7))
sns.countplot(x='sent_bin_3up', data=df, ax=ax, palette='cool_r')
ax.set_title('Distribution of Binary Sentiments for All Apps')
ax.set_ylabel('# of Reviews')
ax.set_xlabel('Sentiment')
ax.set_xticklabels(['Negative (1, 2)', 'Positive (3, 4, 5)'])

for p in ax.patches:
    ax.annotate("%.0f" % p.get_height(), (p.get_x() + p.get_width() / 2.,
                                          p.get_height()),ha='center',
              va='center', fontsize=14, color='black', xytext=(0, 8),
              textcoords='offset points')
fig.tight_layout();
```



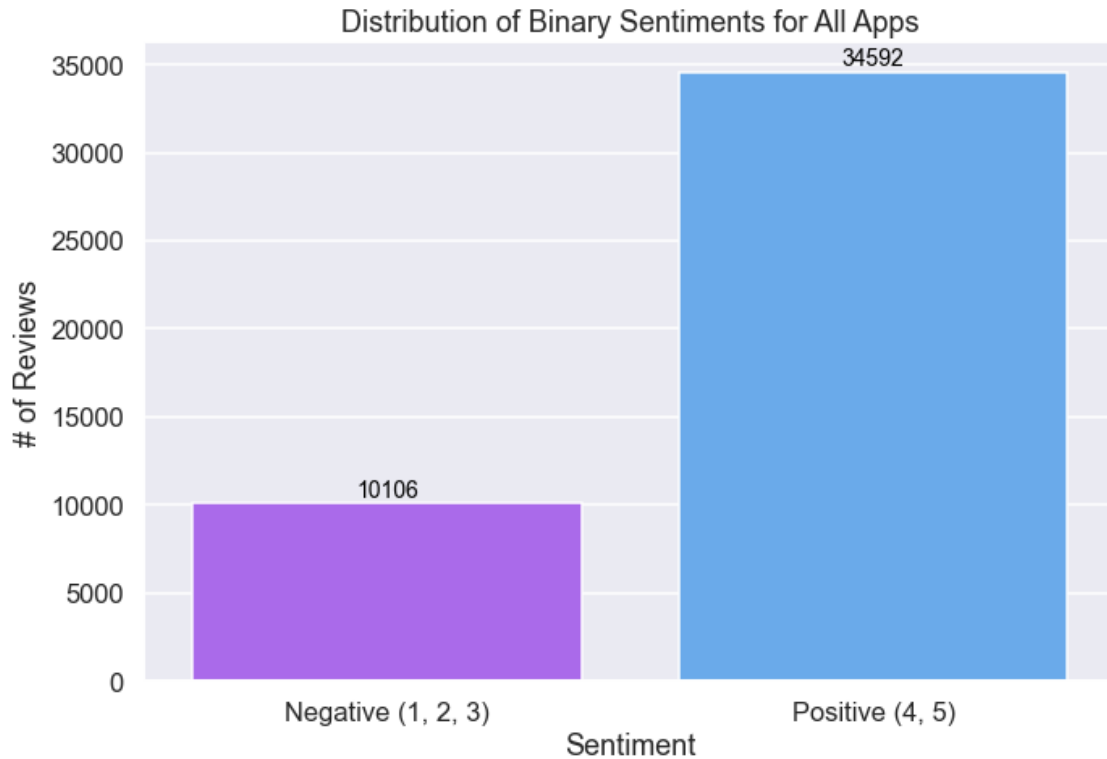
```
[34]: ## Make new binary sentiment column based on ratings 4 and up
conditions = [
    df['rating'] >= 4,
    df['rating'] < 4,
]

values = [1,0]

df['sent_bin_4up'] = np.select(conditions, values)

[35]: # Plot distribution of binary sentiments across all apps
fig, ax = plt.subplots(figsize=(10,7))
sns.countplot(x='sent_bin_4up', data=df, ax=ax, palette='cool_r')
ax.set_title('Distribution of Binary Sentiments for All Apps')
ax.set_ylabel('# of Reviews')
ax.set_xlabel('Sentiment')
ax.set_xticklabels(['Negative (1, 2, 3)', 'Positive (4, 5)'])

for p in ax.patches:
    ax.annotate("%.0f" % p.get_height(), (p.get_x() + p.get_width() / 2.,
                                          p.get_height()),ha='center',
              va='center', fontsize=14, color='black', xytext=(0, 8),
              textcoords='offset points')
fig.tight_layout();
```

We can see a large class imbalance between Negative and Positive for both splits. However, during the modeling trials, it was found that the 1-3, 4-5 split performed slightly better than the 1-2, 3-5 split on all model types. Therefore, we will proceed with the 1-3, 4-5 split, since the insights we can gain from a more accurate model will be more powerful than from a less accurate one.

During the modeling phase, it will be important to keep in mind that there is a class imbalance for both our multi-class classification and our binary classification.

Now, let's slice out the columns relevant to our analysis and move onto preprocessing the actual text reviews.

```
[36]: # Slice out relevant columns into a new dataframe
clean_df = df[['date', 'userName', 'combined_text', 'rating', 'multi_sent',
               'sent_bin_3up', 'sent_bin_4up']]
clean_df
```

```
[36]:
```

	date	userName
0	2021-02-22 18:13:54	Crazy Yorkie Lady
1	2021-02-12 12:42:11	KLC-MHFL
2	2021-02-11 06:22:38	Ashabashley
3	2021-01-18 17:14:21	Kindly38
4	2021-02-08 19:19:10	jlhuggins
...
44717	2020-04-06 23:34:12	JLucret

44718	2020-03-28 06:01:54	mata rato
44719	2020-03-24 13:44:13	Fierce Girl
44720	2020-03-07 17:29:09	Irkesome
44721	2020-02-18 02:08:56	ksavv916

combined_text \

0 This is the BEST app by far for mindfulness. It's almost embarrassing to say this, but I have been a "meditator" for about 30 years and I have to say that I never really got very much out of doing it. I tried so hard, but I'm finding out I never truly understood that "trying" to be a good meditator or mindful person, was missing what doing this was about. What I'm learning through this app is that "mindfulness" is about me - understanding myself first, and it's a lot about acceptance. So with that understanding and acceptance I can begin to see "things" so much more clearly and why previous ways of thinking, particularly about myself, were holding me back from living my life and loving myself. And I'm finding that the more I practice self acceptance, the fears about living fully and going after what I want out of my life, are disappearing. The previous version of myself is rapidly becoming someone I only used to know, and this new ME is much more in line with who I'd always wanted to be. I'm able to try and do so much more, all while in the comfort that I am lovable, even if I fail at something. I simply dust myself off and try some more. This app has completely changed me and the way I'm able to appreciate all the good that's in my life. And I'm worthy of going after the things I want for my life, while freeing me up to share my gifts with others and society in general. If this is where you find your life stalling out, this app will bring you back to yourself.

1 Facing down the rabbit hole. I discovered Headspace on Netflix. My life is a series of invitations to succumb to fear. My husband is a practicing psychologist who is on peritoneal dialysis for 10 hours every night and works remotely with patients for about 10 hours every day without complaint. We have been waiting for a kidney for two years. I am optimistic by nature but even if this was the only focus for concern it would be a lot. Unhelpfully, the list goes on with equivalent issues. I can't work; I need to be on call because I am the go-to for support. I occupy my mind and time writing a book and renovating every room of our home one at a time amidst being available for nine blended (mostly grown) children and 5 grandkids, many of whom have special needs. I am busy but isolated like so many due to covid-19. Trying the Headspace series felt like meeting air; filling the natural and effortless need to breathe. It has provided me with a new lens through which to focus on challenges and see potential. I incorporate it into my waking routine at dawn and it positions me for whatever happens throughout the day, and differentiates a space that is my own. I turn 61 this month and for maybe the first time in a while, life feels lighter and brighter. It's not about what you fear or can't control or even the fact of fear. It's about the lens through which you interpret your fears or challenges and making them your strengths and allies. Thank you Andy :-)

2 A Bright Spot in a Dark World. You won't regret buying

a Headspace subscription. I often joke with people that Headspace is the one thing I would want to take with me to a desert island. If I were to cancel all of my subscriptions today, Headspace is the one exception I would make. I'm a high school English teacher, and our administration put together a month-long mindfulness education program for our students shortly after I subscribed. I had already raved to many of them about the app and the benefits of taking a couple of minutes out of their day to sit and be present. Lo and behold, Andy pops up on the next slide as our Ted Talk speaker of the day! On a more serious note, I suffer from debilitating chronic pain. I'm currently on long-term disability and a leave of absence from the classroom. Mindfulness, and the courses on managing pain, have been a tremendous help in my journey. Not only that, but I struggle with insomnia. My favorite part of Headspace, because of my (admittedly) vivid imagination, are the sleepcasts. I love falling asleep to Rainday Antiques. I can't imagine doing anything but that ever again. I have learned to manage stress and anxiety in a way that I had never thought was possible. I have meditated in the waiting room of The Cleveland Clinic before my appointments. I once laughed at the idea of mindfulness and meditation. Now it's an integral part of my life. It's part of my journey, and I think everyone could benefit from it.

3

The most useful thing I've done in 2020. This app has given me tools to deal with constant change. This past year, I've needed to pivot every day to face some new challenge outside as an essential worker, and at home in isolation with my partner. The tools I learned to use to be present in the moment and to accept change has benefited my partner and colleagues at work as much as my own peace of mind. I have very little control over my environment, but I can see my reaction, accept the emotions that rise up, and move on instead of getting lost or bound by them. This app has also benefited me by helping me get through insomnia. Sometimes it helps me get to sleep in the first place. Mostly it gives me tools to use to calm my mind and body when I wake up in the night buffeted by anxiety or lingering frustration from the day. Lastly, this app has given me a language to express the emotional upheaval and troubling questions that I used to just call anxiety. Naming what makes me feel upset has brought clarity to what can feel like a tempest at the time and allow me to look at my thoughts or, rather, face them. After acknowledging them, I can let them go. Sometimes I'm better at that than at other times. Regular exercise strengthens my focus, my ability to pivot and refocus, and builds roads of calm and quiet that the brain finds easier to fall into and travel due to constant use.

4

It's a Miracle App!. This app has helped me so much! When I am suffering from stress, health anxiety, or fibro/stress related body pains I turn to this app and go through some meditations with the teachers. It's an excellent way for me to relax. Calm my mind. Stop thinking about myself. And to recover in the moment. Hopefully in time this will help me heal, as previously I've just been getting worse and worse without any therapy since my therapist's practice permanently shutdown from COVID. This has been a great help and is something I can just pull out of my pocket and sit in my bedroom for a bit to calm down.

Well worth the 1 year subscription price I've paid. Thank you! *There is only one thing that I'd change. This would be for some of the workout audios. I need to walk a lot for a heart condition I have (which is a big cause of my health anxiety), and I'd love to see the audios for walks and runs to be a bit longer. The longest run audio I see is 20 minutes, when recommended cardio workout lengths realistically should be at least 30 minutes. I'd really like to have an audio similar to the bedtime audios that are lengthy, have soothing talk, and tips to reduce anxiety made to be listened while walking. And if there could be a variety to choose from. Overall this would be the only change I can think of. An hour long walk audio to help calm my mind.

...

...

44717

Me encanta lo mejor para meditar. Es la mejor aplicación se las recomiendo.

44718

Excelente. Lo recomiendo

44719

Love this app. Great app.-would love it if the sounds were organized in a way that makes sense.

44720

Disappointed. You get exactly two meditations before you pay 60.00 per year.

44721

FALSE ADVERTISEMENT subscription terms. As stated in the App Information section WORD FOR WORD: "RelaxMeditation also offers a lifetime subscription for \$19.99 which is paid for by a one-off upfront payment with unlimited access to RelaxMeditation forever." My message to App Support: "I'm trying to purchase the \$19.99 lifetime / unlimited subscription but can't find where / how to?" Their response: "Hello Kristen, thank you for reaching out to us! I really apologize about this. The offers and prices displayed in the Store are not accurate, as they display all of our past and current offers for all of our apps. [...] We do have a special right now, and it offers 60% off the regular price, for 99.99 USD. [...] We also offer a yearly subscription, for 59.99\$, and it includes a 7-day free trial period.[...]" 60% off- at \$99. No thanks. Good luck with that.

	rating	multi_sent	sent_bin_3up	sent_bin_4up
0	5	2	1	1
1	5	2	1	1
2	5	2	1	1
3	5	2	1	1
4	5	2	1	1
...
44717	5	2	1	1
44718	5	2	1	1
44719	4	2	1	1
44720	2	0	0	0
44721	1	0	0	0

```
[44698 rows x 7 columns]
```

5.1 Tokenization

We begin by tokenizing our corpus. Since we do not have any strangely formatted text to worry about with app reviews, we will use the RegExp Tokenizer to tokenize any all words, including those abbreviated with apostrophes.

```
[37]: # Create a single corpus out of all combined text
corpus = clean_df['combined_text'].to_list()

# Convert Corpus to Tokens
tokens = regexp_tokenize(' '.join(corpus), r"([a-zA-Z]+(?:'[a-z]+)?)")

# Preview tokens
tokens[:20]
```

```
[37]: ['This',
      'is',
      'the',
      'BEST',
      'app',
      'by',
      'far',
      'for',
      'mindfulness',
      'It's',
      'almost',
      'embarrassing',
      'to',
      'say',
      'this',
      'but',
      'I',
      'have',
      'been',
      'a']
```

```
[38]: # Check number of unique tokens
len(set(tokens))
```

```
[38]: 28718
```

5.2 Lemmatization

Next, we lemmatize our tokens to make sure that we are able to properly remove any stopwords later.

```
[39]: # Lemmatizing function for list of tokens
def lemmatize_tokens(tokens_list):
    """
    Lemmatizes list of string tokens and returns list of lemmatized
    string tokens.

    Args:
        tokens_list (list) : List of string tokens to be lemmatized.

    Returns:
        tokens_lemm (list) : List of lemmatized string tokens
    """

    lemmatizer = WordNetLemmatizer()
    tokens_lemm = [lemmatizer.lemmatize(word) for word in tokens_list]
    return tokens_lemm
```

```
[40]: # Lemmatize tokens
nltk.download('wordnet')
lemmatized_tokens = lemmatize_tokens(tokens)
```

```
[nltk_data] Downloading package wordnet to
[nltk_data]      /Users/jonathanlee/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
```

```
[41]: # Check number of unique tokens after lemmatization
len(set(lemmatized_tokens))
```

```
[41]: 26678
```

Our lemmatization function has reduced the number of unique tokens by a little over 2000 words.

5.3 Stopword Removal

We can now remove any stopwords that are cluttering our text reviews to prevent us from extracting words that have more value towards determining the user sentiments.

```
[42]: # Write function to create barplot of most common words
def most_freq(tokens, n=25, figsize=(12,7)):
    """
    Displays bar plot of n most frequent words in a list of tokens.

    Args:
        tokens (list) : List of string tokens to plot frequency for.
        n (int) : Number of tokens to plot.
        figsize (float, float) : Width, height in inches.

    Returns:
```

```

"""
# Get frequency distribution of list of tokens
freq = FreqDist(tokens)

# Convert to DataFrame
most_frequent_tokens = pd.DataFrame(freq.most_common(n),
                                     columns=['word', 'count']).sort_values('count',
                                     ↪ascending=False)

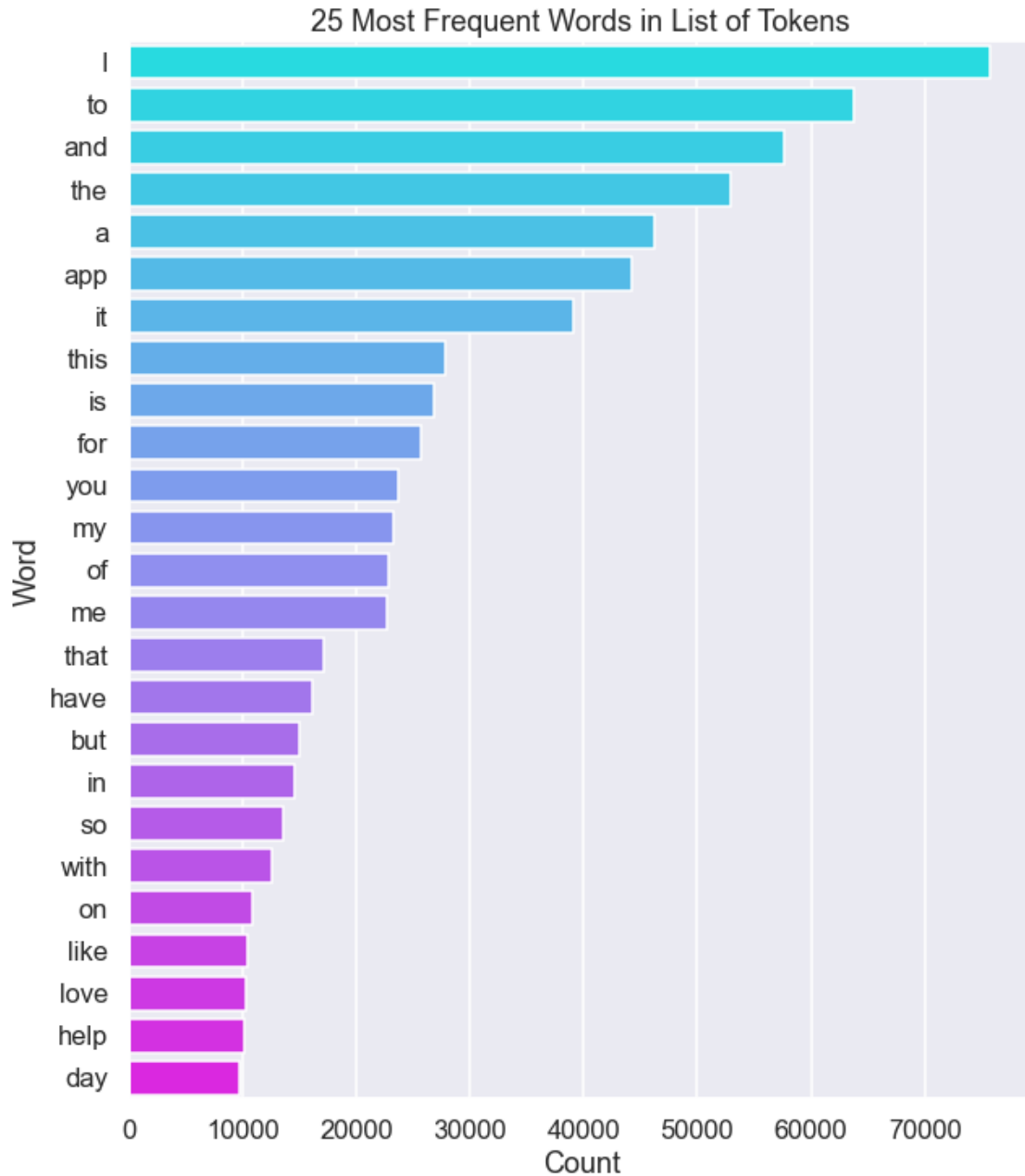
# Plot frequency distribution
fig, ax = plt.subplots(figsize=(10,12))

sns.barplot(data=most_frequent_tokens, y='word', x='count',
            palette='cool', ax=ax, orient='h')

ax.set_title(f'{n} Most Frequent Words in List of Tokens')
ax.set_ylabel('Word')
ax.set_xlabel('Count')

# Display frequency distribution of lemmatized tokens
most_freq(lemmatized_tokens)

```



```
[43]: # Get all the stop words in the English language and preview first 25
      nltk.download('stopwords')
      stopwords_list = stopwords.words('english')
      stopwords_list[:25]
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]   /Users/jonathanlee/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```



```
[43]: ['i',
      'me',
      'my',
      'myself',
      'we',
      'our',
      'ours',
      'ourselves',
      'you',
      "you're",
      "you've",
      "you'll",
      "you'd",
      'your',
      'yours',
      'yourself',
      'yourselves',
      'he',
      'him',
      'his',
      'himself',
      'she',
      "she's",
      'her',
      'hers']
```

```
[44]: # Add punctuation to stopwords_list
stopwords_list.extend(string.punctuation)
stopwords_list[-10:]
```

```
[44]: ['!', '\\\\', ']', '^', '_', '`', '{', '|', '}', '~']
```

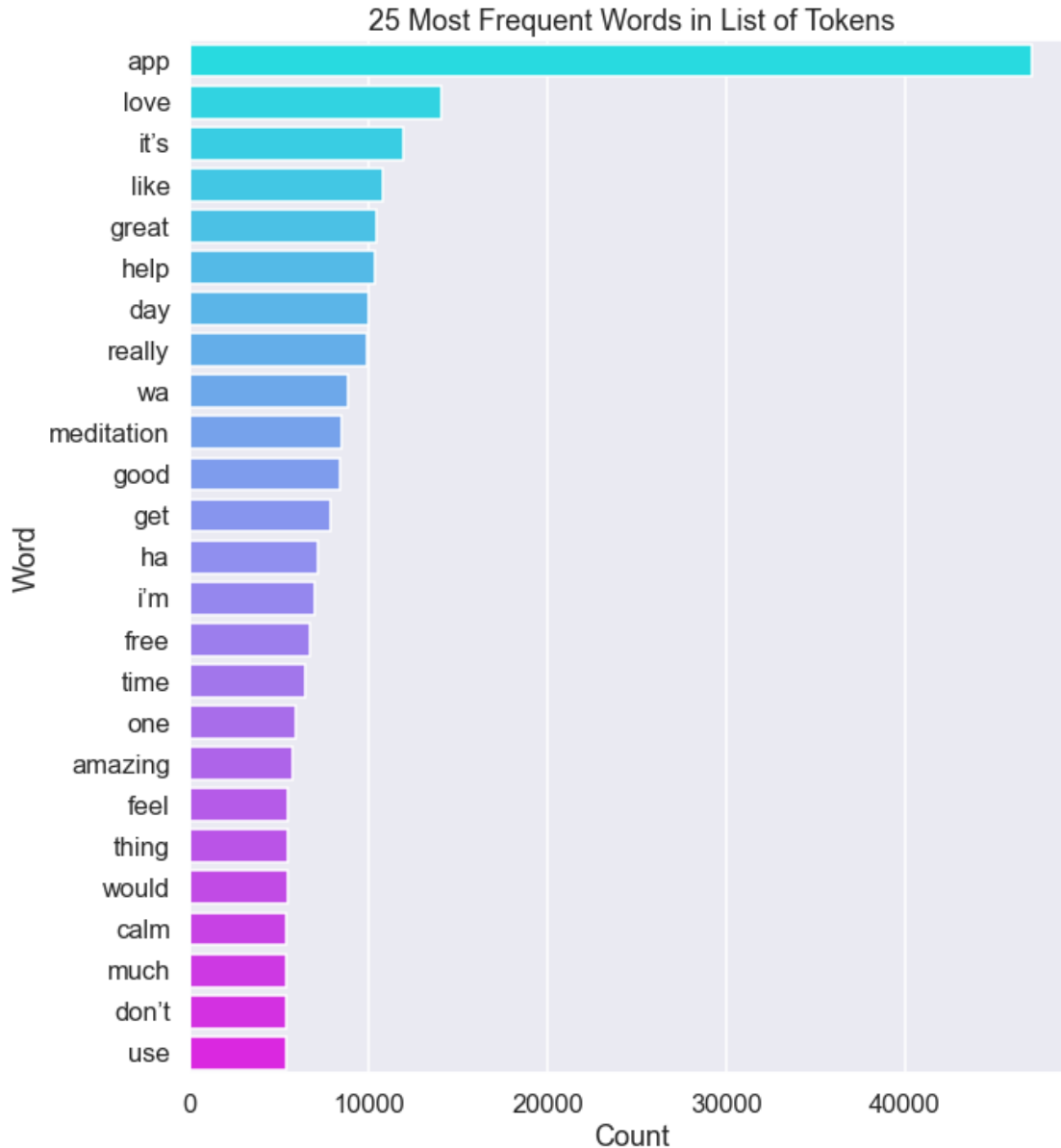
```
[45]: # Add additional punctuation below to stopwords_list
stopwords_list.extend(['"', "'", '...', '"', ',', '~'])
```

```
[46]: # Remove stopwords from list of tokens
stopped_tokens = [w.lower() for w in lemmatized_tokens if w.lower() \
                  not in stopwords_list]
stopped_tokens[:50]
```

```
[46]: ['best',
      'app',
      'far',
      'mindfulness',
      'it's',
      'almost',
      'embarrassing',
```

```
'say',
'meditator',
'year',
'say',
'never',
'really',
'got',
'much',
'tried',
'hard',
'i'm',
'finding',
'never',
'truly',
'understood',
'trying',
'good',
'meditator',
'mindful',
'person',
'wa',
'missing',
'wa',
'i'm',
'learning',
'app',
'mindfulness',
'understanding',
'first',
'it's',
'lot',
'acceptance',
'understanding',
'acceptance',
'begin',
'see',
'thing',
'much',
'clearly',
'previous',
'way',
'thinking',
'particularly']
```

```
[47]: # Display frequency distribution of lemmatized tokens after removing stopwords
most_freq(stopped_tokens, 25, figsize=(12,40))
```



Now that we have removed common stopwords, we begin to see more insightful words. However, we can continue to remove some words that provide no value toward differentiating between negative, positive and neutral sentiments, including the word “app” and the common app names.

```
[48]: # Add common words unrelated to sentiment to stopwords_list
stopwords_list.extend(["app", "wa", "ha", "apps", "headspace", "calm",
                      "insight", "timer", "fabulous", "reflectly"])
```

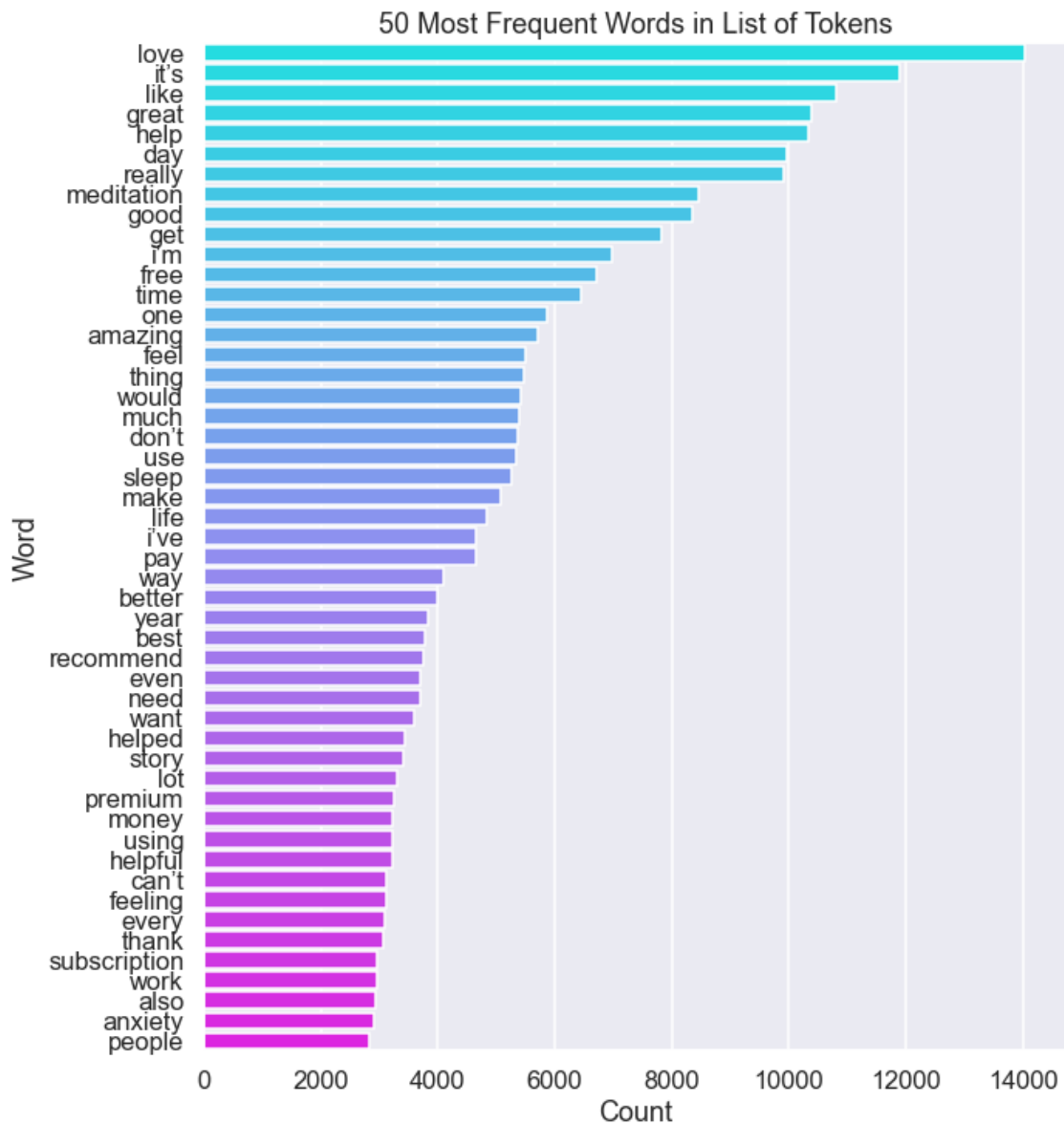
```
[49]: # Remove new stopwords
stopped_tokens = [w.lower() for w in lemmatized_tokens if w.lower() \
                  not in stopwords_list]
```

```
stopped_tokens[:50]
```

```
[49]: ['best',  
      'far',  
      'mindfulness',  
      'it's',  
      'almost',  
      'embarrassing',  
      'say',  
      'meditator',  
      'year',  
      'say',  
      'never',  
      'really',  
      'got',  
      'much',  
      'tried',  
      'hard',  
      'i'm',  
      'finding',  
      'never',  
      'truly',  
      'understood',  
      'trying',  
      'good',  
      'meditator',  
      'mindful',  
      'person',  
      'missing',  
      'i'm',  
      'learning',  
      'mindfulness',  
      'understanding',  
      'first',  
      'it's',  
      'lot',  
      'acceptance',  
      'understanding',  
      'acceptance',  
      'begin',  
      'see',  
      'thing',  
      'much',  
      'clearly',  
      'previous',  
      'way',  
      'thinking',
```

```
'particularly',
'holding',
'back',
'living',
'life']
```

```
[50]: # Display list of tokens after removing extended stopwords
most_freq(stopped_tokens,50, figsize=(12,25))
```



The words that we have left are much more insightful, but we also want to remove words that are too indicative of the user's sentiment. By doing so, we will be able to better examine what elements

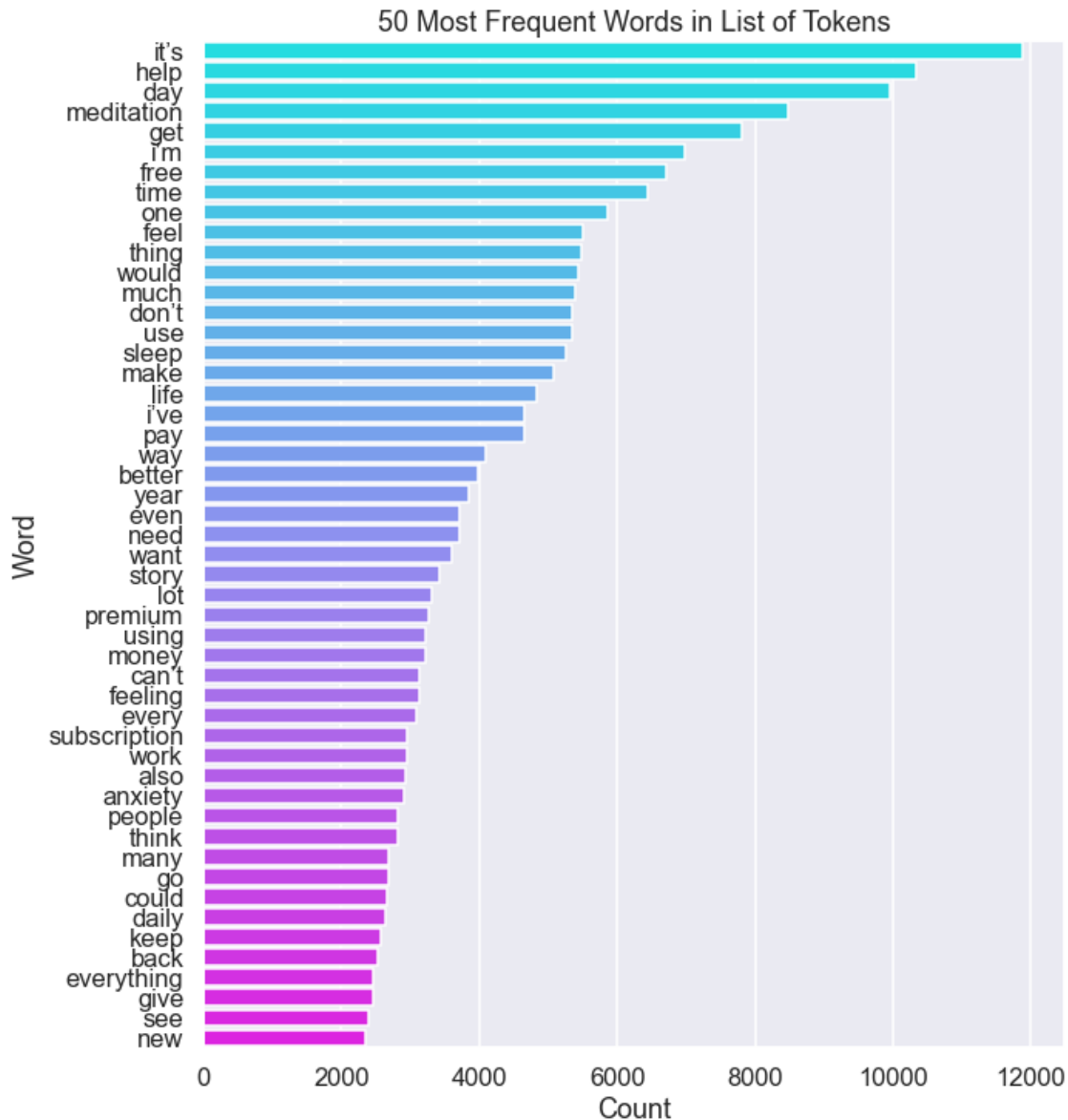
of the apps users considered to be negative or positive without obvious indicator words.

The words that are added to the stopwords list here have been determined after an initial iteration of the modeling process, where obvious words were taking away from our insights regarding app characteristics.

```
[51]: # Add obvious indicator words to stopwords list
stopwords_list.extend(["good", "love", "best", "amazing", "great", "awesome",
                        "perfect", "fabulous", "wonderful", "excellent",
                        "helps", "helped", "helpful", "helping", "loving",
                        "beautiful", "favorite", "fantastic", "calm", "good",
                        "nice", "stupid", "annoying", "hate", "awful", "worse",
                        "worst", "terrible", "disappointed", "disappointing",
                        "garbage", "recommend", "thanks", "thank", "horrible",
                        "like", "it's", "i'm", "really", "like", "bad"])
```

```
[52]: # Remove stopwords
stopped_tokens = [w.lower() for w in lemmatized_tokens if w.lower() \
                  not in stopwords_list]
```

```
[53]: # Display frequency distribution of clean list of tokens
most_freq(stopped_tokens, 50, figsize=(12, 25))
```



5.4 Bigram Exploration

Here, we will briefly look at whether there might be combinations of words that might give us more insight into sentiment than single words.

```
[54]: # Get bigrams and frequencies from list of stopped tokens
bigram_measures = nltk.collocations.BigramAssocMeasures()
bigram_finder = nltk.BigramCollocationFinder.from_words(stopped_tokens)
bigram_scores = bigram_finder.score_ngrams(bigram_measures.raw_freq)
```

```
[55]: # Make a DataFrame from the bigram frequencies
pd.DataFrame(bigram_scores, columns=["Word", "Freq"]).head(20)
```

```
[55]:
```

	Word	Freq
0	(mental, health)	0.001453
1	(fall, asleep)	0.001265
2	(free, trial)	0.001090
3	(sleep, story)	0.001057
4	(make, feel)	0.001053
5	(every, day)	0.000933
6	(guided, meditation)	0.000902
7	(life, changing)	0.000694
8	(easy, use)	0.000680
9	(feel, better)	0.000658
10	(free, version)	0.000649
11	(i've, using)	0.000646
12	(customer, service)	0.000643
13	(help, lot)	0.000623
14	(help, get)	0.000604
15	(don't, want)	0.000596
16	(every, night)	0.000579
17	(don't, know)	0.000536
18	(keep, track)	0.000527
19	(help, sleep)	0.000518

It's clear that there are some quite insightful combinations of words. Sleep seems to be a recurring topic in these bigrams, and we can begin to see that this is one element of a mental health app that we would want to consider allocating resources toward developing. Because bigrams have been seen to be insightful, we will keep note of this as we approach the modeling phase.

Let's move on to examine whether Pointwise Mutual Information scores of our bigrams reveal any insight.

```
[56]: # Get bigrams and PMI scores from list of stopped tokens
bigram_measures = nltk.collocations.BigramAssocMeasures()
bigram_pmi_finder = nltk.BigramCollocationFinder.from_words(stopped_tokens)
bigram_pmi_finder.apply_freq_filter(3)
bigram_pmi_scored = bigram_pmi_finder.score_ngrams(bigram_measures.pmi)
```

```
[57]: # Make a DataFrame from the bigram PMI scores
pd.DataFrame(bigram_pmi_scored, columns=['Words', 'PMI']).head(20)
```

```
[57]:
```

	Words	PMI
0	(mumbo, jumbo)	18.042860
1	(suze, orman)	18.042860
2	(ellie, goulding)	17.627822
3	(gregg, mcbride)	17.627822
4	(idris, elba)	17.627822

5	(jerome, flynn)	17.627822
6	(kabat, zinn)	17.627822
7	(nick, offerman)	17.627822
8	(alice, wonderland)	17.305894
9	(kenneth, soares)	17.305894
10	(roller, coaster)	17.305894
11	(tel, fono)	17.305894
12	(united, states)	17.305894
13	(winnie, pooh)	17.305894
14	(meow, meow)	17.212785
15	(tk, kellman)	17.212785
16	(civil, unrest)	17.042860
17	(sponsored, grubhub)	17.042860
18	(naomi, osaka)	17.042860
19	(marcus, aurelius)	16.890857

Compared to bigram frequencies, PMI scores do not appear to offer much additional insight.

Let's move on to creating wordclouds for our different sentiments.

5.5 WordCloud Visualization

At this point, we have all the parts we need in order to clean up a corpus and create a wordcloud. By creating a wordcloud, we can easily visualize what words are associated with certain sentiments.

Let's begin by previewing a wordcloud for all reviews, with only our final list of stopwords removed.

```
[58]: # Create a WordCloud with our stopwords_list and include bigrams
wordcloud = WordCloud(stopwords=stopwords_list,collocations=True,
                      colormap='Blues', background_color="Black")

# Generate wordcloud from stopped_tokens
wordcloud.generate(' '.join(stopped_tokens))

# Plot with matplotlib
plt.figure(figsize = (12, 12), facecolor = None)
plt.imshow(wordcloud)
plt.axis('off');
```


→
→
→
→
→
→
→
→
→
→

→
→
→
→
→
→
→
→
→
→

→ Meditation should transcend politics. This app has a lot of
→ traditionally based vipassana-style meditation, which is very helpful for
→ learning to meditate and gaining its benefits. However, the app has recently
→ uploaded a lot of political content directly supporting movements such as
→ Black Lives Matter and the ideology of critical race theory. It's hard to
→ emphasize enough how disastrous of a decision this is. It is reducing the
→ transcendent method of meditation to a mere ideological political tool.
→ Meditation allows you the psychological freedom to see passed concepts that
→ you've established unconsciously your whole life and to use that freedom to
→ live deliberately and compassionately with all peoples. Critical race theory
→ attempts to add concepts such as privilege to push you to act in a particular
→ way not necessarily aligned with your deepest intuitions and compassionate
→ motivations. So, it's a reduction of what the practice could be (vipassana,
→ insight, and metta, loving-kindness) for the short-term aims of a political
→ ideology. If you are interested in meditation and not politics, I recommend
→ the Waking Up app, which has vipassana and meta meditations as well as dharma
→ lectures from some eminent practitioners. Good luck on your journey!

→ → → → → → → → → → → →

Doesn't work well with Iwatch. I

↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪

↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪

↪Cannot get a human to respond. I am super frustrated & don't know what else to
↪do to get a human to provide answers. I am having so many problems with the
↪Headspace app- login issues, profile management, invited family members can't
↪create an account. I have tried the FAQs and online support but they don't
↪provide solutions. The "live chat" doesn't exist. To unlink an account it
↪says to contact support - who is that? I have emailed many requests but no
↪answers. 4-6 days to respond due to high volumes. I still haven't heard one
↪iota of anything except that I'm in some queue. I just wanted a human to
↪respond. Nobody has replied but yet my credit card is charged. And on our
↪end, we can't use the app. I realize I am just a drop in the bucket in terms
↪of your users. But I want to cancel my account. I believe I get 30 days. But
↪I'm afraid nobody will respond within that window and I will be stuck with a
↪family subscription nobody can use. Please help me get out of the anxiety that
↪trying to use app has caused. If that ain't irony. What do I need to do to
↪have a customer service representative contact me?

→ → → → → → → → → → → →

U U U U U U U U U U U U

→ → → → → → → → → → → →

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

→completely drains my battery on my iPad 4 when used overnight. Never had this
→issue on other devices.

→ → → → → → → → → →

[illegible]

Hard to cancel □

44703 Was 6 stars. So I got this app around maybe two(ish) years ago and loved it. I suffered from extreme anxiety, agoraphobia and had panic attacks almost every other day. Life was pretty excruciating. After trying the sample meditations I started to notice that my attacks were coming a bit less often. I went ahead and bought the full app (that's right, bought, not rented, not subscribed, bought, for a one time price) as well as talked my wife into trying one of their sleep/clock apps. Everything was great. Over time after meditating around the same time every day, my anxiety almost disappeared. I eventually did most of the meditations multiple times and with my anxiety not quite gone but better than it had been in years, I sat it down for a while while I waited for some new meditations. I'm not gonna lie, I didn't use the app for a long time after that, maybe 6 months or more but it never left my home screen, I never uninstalled it or anything. Eventually my anxiety started bothering me again back around Christmas so I pulled out my handy dandy anxiety busting meditation app only to find that the app that I BOUGHT for just a couple of dollars had been taken back from me and that I could now rent it BACK from them for a ridiculous amount of money a month. My progress is all still there, I'm just locked out of it. I understand the devs have a right to make money and to charge whatever they see fit. I just don't think it's right to sell someone something only to take it back away from them and try to rent it back out to them again for thousands of more dollars. If it was to be changed I feel like the changes (a monthly subscription fee) should have only applied to people who downloaded/installed the app AFTER the decision to change it from a flat price to a subscription fee had been made and not forced upon those who had already paid for the full version. I was originally gonna leave a review around Christmas when I discovered what had happened but was just blown away and I guess hurt. I didn't want to talk or even think of it. Smh

[illegible]

[illegible]

39	2021-05-06 17:27:37	whatnickname256
...
44579	2020-02-22 18:10:33	Lithmac
44589	2020-04-04 19:42:41	
44602	2020-02-14 03:52:19	Papa's apple
44624	2020-05-26 19:55:30	J4m3\$rawdogger
44671	2020-06-17 03:12:20	the real body of rage

→ combined_text \

10 Great resource!. This is really helpful! The only
→ reason I didn't give a higher rating is because there's a strong overlying
→ tone of a sales pitch there. Which is unfortunate, because I think people
→ would be able to see the value easier if there wasn't a push for money after
→ every meditation and for many of the things on the app people would want to
→ explore. Kind of stalls progress of the meditation itself when you've just
→ worked so hard to pull yourself back, only to be instantly asked for money
→ right after. From a company stand point I also think it would be more
→ beneficial to work on building value and then freely offering the option to
→ see what monthly costs are and also show a detailed description of what's
→ involved with a paid membership. I know this has worked well financially for
→ other services and I've seen increases in sales by simply looking at things
→ from the customers perspective. And I feel like the customers perspective is a
→ miss here :(However quite and easy fix and overall this app is amazing! It
→ not only teaches the basics of meditation, but teaches different techniques as
→ well. This makes it good for beginners all the way to experts. I truly believe
→ that the basis of this app is pure in intention and I can see what a help it
→ is to many people who walk the path of life.

16 A great app. Needs work. I give this app 5 stars because the build of
→this app, the contents of this app, the flow of this app is great. well done.
→Every app can be improved and this one needs work in some ways 1. the watch
→app: the only thing I every get on the watch app is a 1 minute "breath mini".
→the watch app is basically useless. Why can't I see my today screen on the
→watch and start my various meditations from the watch. the watch app gets 1
→star. 2. Shortcuts: on the iphone I have setup a sleep schedule. On the "good
→night" screen I am given the option to run shortcuts. Headspace offers a few
→shortcuts that can be called from this screen. One is the nightly "sleepcast".
→the sleepcast is supposed to be different every night. It is different on my
→today screen, but the shortcut always plays the "Rainyday Antiques". therefore
→the shortcut is useless. this part of the app gets 1 star. Lastly, the
→philosophy behind this app is mindlessness. It is taught as a universal fix
→for bringing the mind to a place of ease. As such it is a religious philosophy.
→ I think they need to be more open about this and they need to state that the
→point of this is to alleviate suffering. This app looks like they did version
→1 then stopped before finishing the details. Shortcuts must work for this
→app to be fully functional

23

→
→
→ Ok but could be better. personally, i love this app. the
→meditations are really helpful for my crippling anxiety, and i thoroughly
→enjoy watching the daily videos and doing the mindfulness moments. i also
→really like the new updated version with the dark mode and everything.
→however, i'm not as happy with headspace as i thought i would be when i got
→the app. i'm still a kid, and i'm not going to spend money to get headspace+,
→so i've had the normal version, however i'm very disappointed about the lack
→of content i have unlocked. i can only do a few meditations and they're
→limited. this is very upsetting to me, and i've heard that others have felt
→the same way. i'm considering deleting headspace for exactly this reason:
→there's not enough free content. soon enough i will run out of meditation
→sessions. i am not satisfied with this app. PLEASE unlock more content for
→those who don't want or CANT spend money on headspace+!!!! you might lose this
→costumer to your app for just that reason. i need more content because my
→anxiety lately has been through the roof, so i would be excited to try out
→some new meditations but i CANT. please fix this!!!!

26

↳
↳
↳ Was hoping for more. Not
↳ every technique works equally well for everyone, and I feel like they put an
↳ over emphasis on the body scan (which is admittedly one I'm not particularly
↳ fond of). They have courses like "dealing with anxiety" or "managing anger",
↳ but they're all essentially the same: "here's some basic advice you'll get
↳ from every therapist for 30 seconds, now let's do 20 minutes of body scan." I
↳ don't feel like they really apply the meditation to the issues often enough,
↳ but the advice is decent. Also, in the beginners courses they mention a lot
↳ of abstract concepts that are supposed to help you but they don't explain them
↳ very well or reinforce them very often. So half the time I find myself
↳ thinking "Wait, what does he mean by that again and how was I supposed go
↳ about it?" There are enough great meditation pointers though that it's been
↳ super helpful to get over some hurdles I was struggling with and the voices
↳ are very relaxing. Unfortunately the content is a bit sparse given how much
↳ stuff they have posted and it can be difficult finding the gems when there's
↳ so much filler to wade through.

39

↳
↳
↳
↳ Please stop those
↳ anxiety-triggering pop-up messages!. I am having my COVID shot the next day
↳ and REALLY wanted to have a nice sleep before that. I took one meditation
↳ session before going to bed in headspace. As usual, it worked nicely BUT when
↳ I reached to my phone to close the app, a pop up message showed up saying
↳ "Can't sleep? Try our new sleepcast" it immediately triggered my anxiety of
↳ afraid of losing sleep, and I wasn't able to sleep that night and had to
↳ cancel my COVID shot appointment. This is really disappointing from a
↳ meditation app as this kind of message can easily trigger someone who's
↳ already anxious (and those are often the users of headspace). Especially this
↳ is an in-app pop up that you cannot disable in the iOS system setting. I am
↳ not sure why this message will show up at that time and I don't want to assume
↳ bad intentions (after all Headspace had helped me quite a lot and I am
↳ thankful). Please stop sending these kinds of anxiety-triggering messages or
↳ at least allow us to disable them! From a 5-year customer.

...

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

44579

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

Buggy music. Decent

↳ white noise generator, but the feature that claims to play music from the
↳ music library is hit and miss and miss and miss. I don't have any iCloud music-
everything is downloaded to local storage- but the app grays out about 3/4 of my
↳ tracks claiming they're iCloud. Naturally, all the tracks I'd want to use with
↳ this get grayed out. After those bugs get fixed, it might be worth its price.

44589

↪

↪

↪

↪

↪

↪

↪

↪

↪

↪

↪

↪

↪

Good app but why is the timer gone. Why is the
↪timer for turning off the music gone? This is annoying because I went to
↪sleep, woke up and went downstairs for a while, then when I go upstairs the
↪music is still playing! Luckily it doesn't use too much battery life, but
↪still!

44602

↪

↪

↪

↪

↪

↪

↪

↪

↪

↪

↪

↪

↪

Ahhhh sleeeeeeep . I use this app every
↪night too (as I see others comment the same)! It sure calms my thoughts and
↪gives me a relaxing focus! A few of the sounds have irritating repetitive
↪background noises or fluctuations. I love the thunder and slow waves the most!

44624

↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪

␣
␣
␣
␣
␣
␣
␣
␣
␣
␣
␣
␣
␣
␣

↪Great so far. Really liking the app so far. Would like more variety of
↪meditations and sounds though. FYI This app advertises a lifetime membership
↪for \$20 and they don't actually offer that deal, so that is extremely
↪misleading!

44671

↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪
↪

␣
␣
␣
␣
␣
␣
␣
␣
␣
␣
␣
␣
␣
␣
␣
␣
␣

↪Not bad. I like the gift
↪and thanks you it's really help but the price is still a bit considering

	rating	multi_sent	sent_bin_3up	sent_bin_4up
10	3	1	1	0
16	3	1	1	0
23	3	1	1	0
26	3	1	1	0
39	3	1	1	0
...
44579	3	1	1	0
44589	3	1	1	0
44602	3	1	1	0
44624	3	1	1	0

44671 3 1 1 0

[2379 rows x 7 columns]

		date	userName \
0	2021-02-22 18:13:54	Crazy Yorkie Lady	
1	2021-02-12 12:42:11	KLC-MHFL	
2	2021-02-11 06:22:38	Ashabashley	
3	2021-01-18 17:14:21	Kindly38	
4	2021-02-08 19:19:10	jlhuggins	
...	
44715	2020-05-05 14:01:25	Randy191987	
44716	2020-05-04 21:18:00	Mlpk182	
44717	2020-04-06 23:34:12	JLucret	
44718	2020-03-28 06:01:54	mata rato	
44719	2020-03-24 13:44:13	Fierce Girl	

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳ \

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

↳

combined_text ↳

2

A Bright Spot in a Dark World. You won't regret buying
↳ a Headspace subscription. I often joke with people that Headspace is the one
↳ thing I would want to take with me to a desert island. If I were to cancel all
↳ of my subscriptions today, Headspace is the one exception I would make. I'm a
↳ high school English teacher, and our administration put together a month-long
↳ mindfulness education program for our students shortly after I subscribed. I
↳ had already raved to many of them about the app and the benefits of taking a
↳ couple of minutes out of their day to sit and be present. Lo and behold, Andy
↳ pops up on the next slide as our Ted Talk speaker of the day! On a more
↳ serious note, I suffer from debilitating chronic pain. I'm currently on
↳ long-term disability and a leave of absence from the classroom. Mindfulness,
↳ and the courses on managing pain, have been a tremendous help in my journey.
↳ Not only that, but I struggle with insomnia. My favorite part of Headspace,
↳ because of my (admittedly) vivid imagination, are the sleepcasts. I love
↳ falling asleep to Rainday Antiques. I can't imagine doing anything but that
↳ ever again. I have learned to manage stress and anxiety in a way that I had
↳ never thought was possible. I have meditated in the waiting room of The
↳ Cleveland Clinic before my appointments. I once laughed at the idea of
↳ mindfulness and meditation. Now it's an integral part of my life. It's part of
↳ my journey, and I think everyone could benefit from it.

3

↳ The most useful thing I've done in 2020. This app has
↳ given me tools to deal with constant change. This past year, I've needed to
↳ pivot every day to face some new challenge outside as an essential worker, and
↳ at home in isolation with my partner. The tools I learned to use to be present
↳ in the moment and to accept change has benefited my partner and colleagues at
↳ work as much as my own peace of mind. I have very little control over my
↳ environment, but I can see my reaction, accept the emotions that rise up, and
↳ move on instead of getting lost or bound by them. This app has also benefited
↳ me by helping me get through insomnia. Sometimes it helps me get to sleep in
↳ the first place. Mostly it gives me tools to use to calm my mind and body when
↳ I wake up in the night buffeted by anxiety or lingering frustration from the
↳ day. Lastly, this app has given me a language to express the emotional
↳ upheaval and troubling questions that I used to just call anxiety. Naming what
↳ makes me feel upset has brought clarity to what can feel like a tempest at the
↳ time and allow me to look at my thoughts or, rather, face them. After
↳ acknowledging them, I can let them go. Sometimes I'm better at that than at
↳ other times. Regular exercise strengthens my focus, my ability to pivot and
↳ refocus, and builds roads of calm and quiet that the brain finds easier to
↳ fall into and travel due to constant use.

α_L

U

L

●●●

44715

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

Excelente. Excelente aplicación es muy buena

44716

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

Excelente. Excelente para dormir meditar o concentrarse la amo !

44717

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

Me encanta lo mejor para meditar. Es la mejor aplicación se las recomiendo.

44718

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

Excelente. Lo recomiendo

U
U
U
U
U
U
U
U
U
U
U
U
U
U
U

Great resource!. This is really helpful! The only reason I didn't give a higher rating is because there's a strong overlying tone of a sales pitch there. Which is unfortunate, because I think people would be able to see the value easier if there wasn't a push for money after every meditation and for many of the things on the app people would want to explore. Kind of stalls progress of the meditation itself when you've just worked so hard to pull yourself back, only to be instantly asked for money right after. From a company stand point I also think it would be more beneficial to work on building value and then freely offering the option to see what monthly costs are and also show a detailed description of what's involved with a paid membership. I know this has worked well financially for other services and I've seen increases in sales by simply looking at things from the customers perspective. And I feel like the customers perspective is a miss here :(However quite and easy fix and overall this app is amazing! It not only teaches the basics of meditation, but teaches different techniques as well. This makes it good for beginners all the way to experts. I truly believe that the basis of this app is pure in intention and I can see what a help it is to many people who walk the path of life.

→
→
→
→
→
→
→
→
→

→
→
→
→
→
→
→
→
→

→ A great app. Needs work. I give this app 5 stars. →
→ because the build of this app, the contents of this app, the flow of this app →
→ is great. well done. Every app can be improved and this one needs work in some →
→ ways 1. the watch app: the only thing I every get on the watch app is a 1 →
→ minute "breath mini". the watch app is basically useless. Why can't I see my →
→ today screen on the watch and start my various meditations from the watch. the →
→ watch app gets 1 star. 2. Shortcuts: on the iphone I have setup a sleep →
→ schedule. On the "good night" screen I am given the option to run shortcuts. →
→ Headspace offers a few shortcuts that can be called from this screen. One is →
→ the nightly "sleepcast". the sleepcast is supposzed to be different every →
→ night. It is different on my today screen, but the shortcut always plays the →
→ "Rainyday Antiques". therefore the shortcut is useless. this part of the app →
→ gets 1 star. Lastly, the philosophy behind this app is mindlessness. It is →
→ taught as a universal fix for bringing the mind to a place of ease. As such it →
→ is a religious philosophy. I think they need to be more open about this and →
→ they need to state that the point of this is to alleviate suffering. This app →
→ looks like they did version 1 then stopped before finishing the details. →
→ Shortcuts must work for this app to be fully functional

↳
↳
↳
↳
↳
↳
↳
↳
↳
↳
↳

↳
↳
↳
↳
↳
↳
↳
↳
↳
↳
↳

Ok but could be better. personally, i
↳love this app. the meditations are really helpful for my crippling anxiety,↳
↳and i thoroughly enjoy watching the daily videos and doing the mindfulness↳
↳moments. i also really like the new updated version with the dark mode and↳
↳everything. however, i'm not as happy with headspace as i thought i would be↳
↳when i got the app. i'm still a kid, and i'm not going to spend money to get↳
↳headspace+, so i've had the normal version, however i'm very disappointed↳
↳about the lack of content i have unlocked. i can only do a few meditations and↳
↳they're limited. this is very upsetting to me, and i've heard that others have↳
↳felt the same way. i'm considering deleting headspace for exactly this reason:↳
↳there's not enough free content. soon enough i will run out of meditation↳
↳sessions. i am not satisfied with this app. PLEASE unlock more content for↳
↳those who don't want or CANT spend money on headspace+!!!! you might lose this↳
↳costumer to your app for just that reason. i need more content because my↳
↳anxiety lately has been through the roof, so i would be excited to try out↳
↳some new meditations but i CANT. please fix this!!!!

Was_U

88

Please stop.

...

...

[illegible]

☐ ☐

→completely drains my battery on my iPad 4 when used overnight. Never had this
→issue on other devices.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Hard to cancel ▢

44703 Was 6 stars. So I got this app around maybe two(ish) years ago and loved it. I suffered from extreme anxiety, agoraphobia and had panic attacks almost every other day. Life was pretty excruciating. After trying the sample meditations I started to notice that my attacks were coming a bit less often. I went ahead and bought the full app (that's right, bought, not rented, not subscribed, bought, for a one time price) as well as talked my wife into trying one of their sleep/clock apps. Everything was great. Over time after meditating around the same time every day, my anxiety almost disappeared. I eventually did most of the meditations multiple times and with my anxiety not quite gone but better than it had been in years, I sat it down for a while while I waited for some new meditations. I'm not gonna lie, I didn't use the app for a long time after that, maybe 6 months or more but it never left my home screen, I never uninstalled it or anything. Eventually my anxiety started bothering me again back around Christmas so I pulled out my handy dandy anxiety busting meditation app only to find that the app that I BOUGHT for just a couple of dollars had been taken back from me and that I could now rent it BACK from them for a ridiculous amount of money a month. My progress is all still there, I'm just locked out of it. I understand the devs have a right to make money and to charge whatever they see fit. I just don't think it's right to sell someone something only to take it back away from them and try to rent it back out to them again for thousands of more dollars. If it was to be changed I feel like the changes (a monthly subscription fee) should have only applied to people who downloaded/installed the app AFTER the decision to change it from a flat price to a subscription fee had been made and not forced upon those who had already paid for the full version. I was originally gonna leave a review around Christmas when I discovered what had happened but was just blown away and I guess hurt. I didn't want to talk or even think of it. Smh

[illegible]

U

U U U U U U U U U U U U U U U

m_{d}
 d_{r}
 r_{r}
 r_{d}
 ρ_{d}

4	2021-02-08 19:19:10	jlhuggins
...
44715	2020-05-05 14:01:25	Randy191987
44716	2020-05-04 21:18:00	Mlpk182
44717	2020-04-06 23:34:12	JLucret
44718	2020-03-28 06:01:54	mata rato
44719	2020-03-24 13:44:13	Fierce Girl

0 This is the BEST app by far for mindfulness. It's almost embarrassing to
→say this, but I have been a "meditator" for about 30 years and I have to say
→that I never really got very much out of doing it. I tried so hard, but I'm
→finding out I never truly understood that "trying" to be a good meditator or
→mindful person, was missing what doing this was about. What I'm learning
→through this app is that "mindfulness" is about me - understanding myself
→first, and it's a lot about acceptance. So with that understanding and
→acceptance I can begin to see "things" so much more clearly and why previous
→ways of thinking, particularly about myself, were holding me back from living
→my life and loving myself. And I'm finding that the more I practice self
→acceptance, the fears about living fully and going after what I want out of my
→life, are disappearing. The previous version of myself is rapidly becoming
→someone I only used to know, and this new ME is much more in line with who I'd
→always wanted to be. I'm able to try and do so much more, all while in the
→comfort that I am lovable, even if I fail at something. I simply dust myself
→off and try some more. This app has completely changed me and the way I'm able
→to appreciate all the good that's in my life. And I'm worthy of going after
→the things I want for my life, while freeing me up to share my gifts with
→others and society in general. If this is where you find your life stalling
→out, this app will bring you back to yourself.

1

Facing down the rabbit hole. I discovered

↳ Headspace on Netflix. My life is a series of invitations to succumb to fear.↳
↳ My husband is a practicing psychologist who is on peritoneal dialysis for 10↳
↳ hours every night and works remotely with patients for about 10 hours every↳
↳ day without complaint. We have been waiting for a kidney for two years. I am↳
↳ optimistic by nature but even if this was the only focus for concern it would↳
↳ be a lot. Unhelpfully, the list goes on with equivalent issues. I can't work;↳
↳ I need to be on call because I am the go-to for support. I occupy my mind and↳
↳ time writing a book and renovating every room of our home one at a time amidst↳
↳ being available for nine blended (mostly grown) children and 5 grandkids, many↳
↳ of whom have special needs. I am busy but isolated like so many due to↳
↳ covid-19. Trying the Headspace series felt like meeting air; filling the↳
↳ natural and effortless need to breathe. It has provided me with a new lens↳
↳ through which to focus on challenges and see potential. I incorporate it into↳
↳ my waking routine at dawn and it positions me for whatever happens throughout↳
↳ the day, and differentiates a space that is my own. I turn 61 this month and↳
↳ for maybe the first time in a while, life feels lighter and brighter. It's not↳
↳ about what you fear or can't control or even the fact of fear. It's about the↳
↳ lens through which you interpret your fears or challenges and making them your↳
↳ strengths and allies. Thank you Andy :-)

2

A Bright Spot in a Dark World. You won't regret buying

↳ a Headspace subscription. I often joke with people that Headspace is the one↳
↳ thing I would want to take with me to a desert island. If I were to cancel all↳
↳ of my subscriptions today, Headspace is the one exception I would make. I'm a↳
↳ high school English teacher, and our administration put together a month-long↳
↳ mindfulness education program for our students shortly after I subscribed. I↳
↳ had already raved to many of them about the app and the benefits of taking a↳
↳ couple of minutes out of their day to sit and be present. Lo and behold, Andy↳
↳ pops up on the next slide as our Ted Talk speaker of the day! On a more↳
↳ serious note, I suffer from debilitating chronic pain. I'm currently on↳
↳ long-term disability and a leave of absence from the classroom. Mindfulness,↳
↳ and the courses on managing pain, have been a tremendous help in my journey.↳
↳ Not only that, but I struggle with insomnia. My favorite part of Headspace,↳
↳ because of my (admittedly) vivid imagination, are the sleepcasts. I love↳
↳ falling asleep to Rainday Antiques. I can't imagine doing anything but that↳
↳ ever again. I have learned to manage stress and anxiety in a way that I had↳
↳ never thought was possible. I have meditated in the waiting room of The↳
↳ Cleveland Clinic before my appointments. I once laughed at the idea of↳
↳ mindfulness and meditation. Now it's an integral part of my life. It's part of↳
↳ my journey, and I think everyone could benefit from it.

3

→ The most useful thing I've done in 2020. This app has
→ given me tools to deal with constant change. This past year, I've needed to
→ pivot every day to face some new challenge outside as an essential worker, and
→ at home in isolation with my partner. The tools I learned to use to be present
→ in the moment and to accept change has benefited my partner and colleagues at
→ work as much as my own peace of mind. I have very little control over my
→ environment, but I can see my reaction, accept the emotions that rise up, and
→ move on instead of getting lost or bound by them. This app has also benefited
→ me by helping me get through insomnia. Sometimes it helps me get to sleep in
→ the first place. Mostly it gives me tools to use to calm my mind and body when
→ I wake up in the night buffeted by anxiety or lingering frustration from the
→ day. Lastly, this app has given me a language to express the emotional
→ upheaval and troubling questions that I used to just call anxiety. Naming what
→ makes me feel upset has brought clarity to what can feel like a tempest at the
→ time and allow me to look at my thoughts or, rather, face them. After
→ acknowledging them, I can let them go. Sometimes I'm better at that than at
→ other times. Regular exercise strengthens my focus, my ability to pivot and
→ refocus, and builds roads of calm and quiet that the brain finds easier to
→ fall into and travel due to constant use.

4

→ It's a Miracle App!. This app has helped
→ me so much! When I am suffering from stress, health anxiety, or fibro/stress
→ related body pains I turn to this app and go through some meditations with the
→ teachers. It's an excellent way for me to relax. Calm my mind. Stop thinking
→ about myself. And to recover in the moment. Hopefully in time this will help
→ me heal, as previously I've just been getting worse and worse without any
→ therapy since my therapist's practice permanently shutdown from COVID. This
→ has been a great help and is something I can just pull out of my pocket and
→ sit in my bedroom for a bit to calm down. Well worth the 1 year subscription
→ price I've paid. Thank you! *There is only one thing that I'd change. This
→ would be for some of the workout audios. I need to walk a lot for a heart
→ condition I have (which is a big cause of my health anxiety), and I'd love to
→ see the audios for walks and runs to be a bit longer. The longest run audio I
→ see is 20 minutes, when recommended cardio workout lengths realistically
→ should be at least 30 minutes. I'd really like to have an audio similar to the
→ bedtime audios that are lengthy, have soothing talk, and tips to reduce
→ anxiety made to be listened while walking. And if there could be a variety to
→ choose from. Overall this would be the only change I can think of. An hour
→ long walk audio to help calm my mind.

[illegible]

Excelente. Excelente aplicación es muy buena

44716

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

Excelente. Excelente para dormir meditar o concentrarse la amo !

44717

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

↩

Me encanta lo mejor para meditar. Es la mejor aplicación se las recomiendo.

U U U U U U U U U U U U U U U U U

[illegible][illegible]

44715	5	2	1	1
44716	5	2	1	1
44717	5	2	1	1
44718	5	2	1	1
44719	4	2	1	1

[34592 rows x 7 columns]

```
[61]: # Get image file to use for wordcloud mask
urllib.request.urlretrieve(
    'https://github.com/ds-leeahanjin/app-reviews-nlp-sentiment-analysis/blob/
    ↪master/images/meditating_girl.png?raw=true',
    "meditating_girl.png")

# Create wordcloud mask
mask = np.array(Image.open('meditating_girl.png'))
```

```
[62]: # Write functions to use to create wordclouds
def wordcloud_prep(df_col, stopwords):
    """
    Converts a column of strings into a list of lemmatized tokens with
    a specified list of stopwords removed.

    Args:
        df_col (col) : Sliced column from dataframe with text to be put into
                        a wordcloud
        stopwords (list) : List of stopwords to remove from text.

    Returns:
        stopped_cloud (list) : List of lemmatized tokens with stopwords
                              removed.
    """

    # Tokenize and lemmatize corpus into tokens
    cloud_corpus = df_col.to_list()
    cloud_tokens = rexp_tokenize(' '.join(cloud_corpus),
                                   r"([a-zA-Z]+(?:'[a-z]+)?)")
    lemmatized_cloud = lemmatize_tokens(cloud_tokens)

    # Remove specified stopwords
    stopped_cloud = [w.lower() for w in lemmatized_cloud if w.lower() \
                     not in stopwords]
    return stopped_cloud

def generate_wordcloud(tokens, stopwords=None, collocations=False,
                       background_color='black', colormap='cool',
```

```

        contour_color='steelblue',
        mask=mask):
    """
    Creates and displays a wordcloud using the specified list of tokens,
    stopwords, and mask image.

    Args:
        tokens (list) : List of tokens to generate wordcloud from.
        stopwords (list) : List of stopwords to remove from text.
        collocations (bool) : Whether to include collocations (bigrams) of
                             two words.
        background_color (color value) : Background color for the wordcloud_
→image.
        colormap (matplotlib colormap) : Matplotlib colormap to randomly draw
                                         colors from for each word.
        contour_color : Mask contour color.
        mask (nd-array) : Gives a binary mask on where to draw words. If mask
                          is not None, width and height will be ignored and
                          the shape of mask will be used instead.

    Returns:
        wordcloud : Image of wordcloud generated using specified text.
    """

    # Initialize a WordCloud
    wordcloud = WordCloud(stopwords = stopwords,
                          collocations=collocations,
                          background_color=background_color,
                          colormap=colormap,
                          mask=mask, contour_width=2,
                          contour_color=contour_color,
                          min_font_size=15)

    # Generate wordcloud from tokens
    wordcloud.generate(','.join(tokens))

    plt.figure(figsize = (12, 12), facecolor = None)
    plt.imshow(wordcloud)
    plt.axis('off');
    return wordcloud

```

5.5.2 Multi-Class Sentiment WordClouds

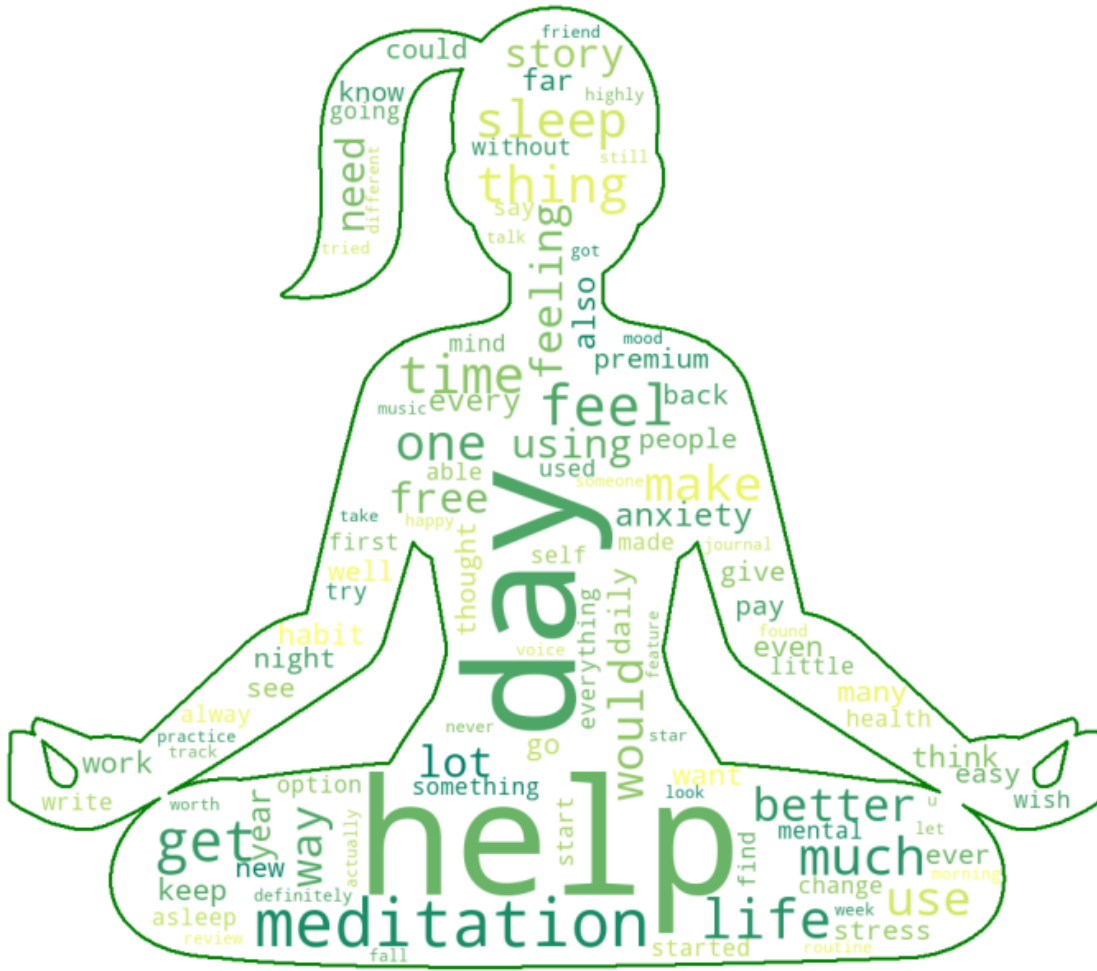
Now, let's put everything together to generate some wordclouds on our multi-class sentiments.

```

[63]: # Generate positive wordcloud for multi-class sentiment
multi_pos_tokens = wordcloud_prep(multi_pos_df['combined_text'], stopwords_list)

```

```
multi_pos_wc = generate_wordcloud(multi_pos_tokens, stopwords_list,
                                  colormap='summer', background_color='white',
                                  contour_color='green')
```



```
[64]: # Generate negative wordcloud for multi-class sentiment
multi_neg_tokens = wordcloud_prep(multi_neg_df['combined_text'], stopwords_list)
multi_neg_wc = generate_wordcloud(multi_neg_tokens, stopwords_list,
                                  colormap='YlOrRd', background_color='white',
                                  contour_color='red')
```

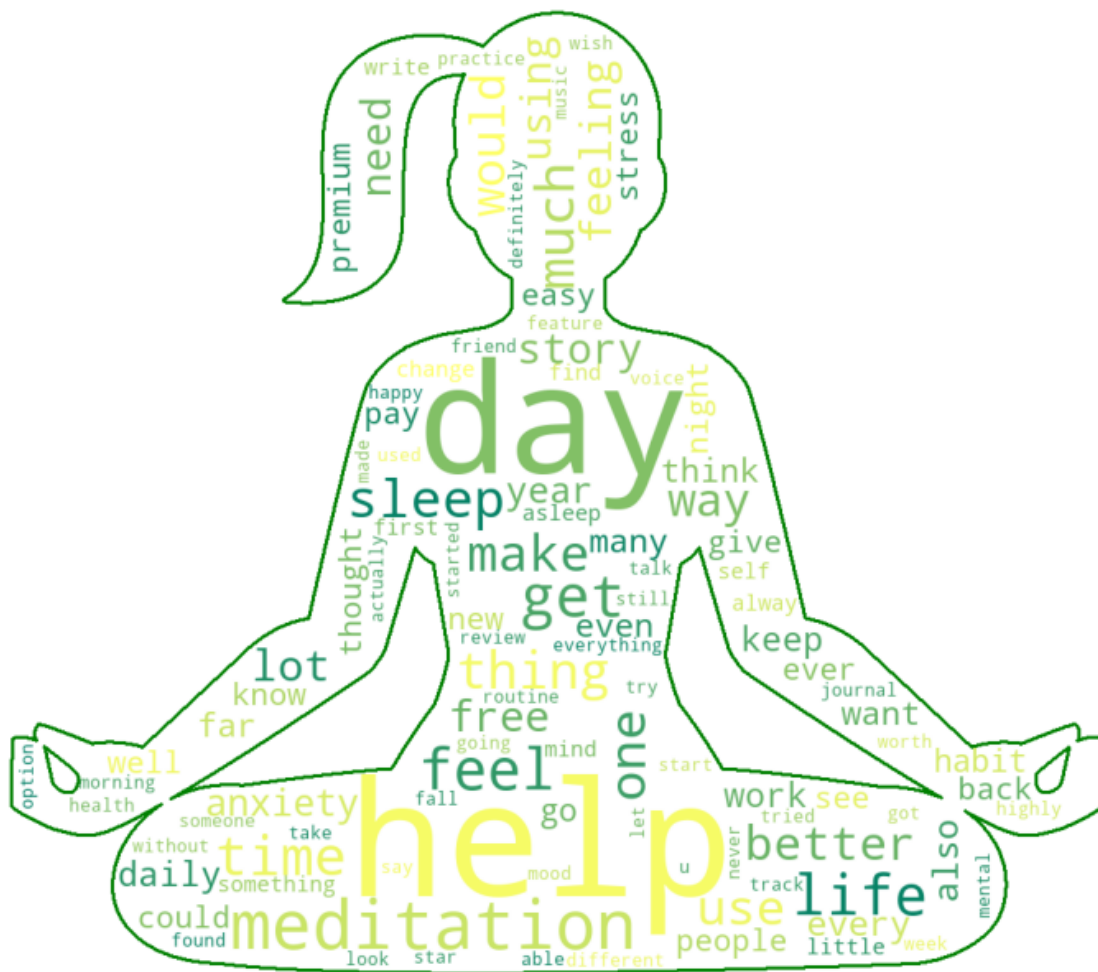



The positive and neutral wordclouds that we see above indicate that there are a variety of topics that users write about in their reviews. Note that we can see sleep again seems to be one of the more important topics for neutral and positive reviews. However, it's much more obvious for the negative wordcloud that the majority of dissatisfaction that users have is either payment or customer-support related.

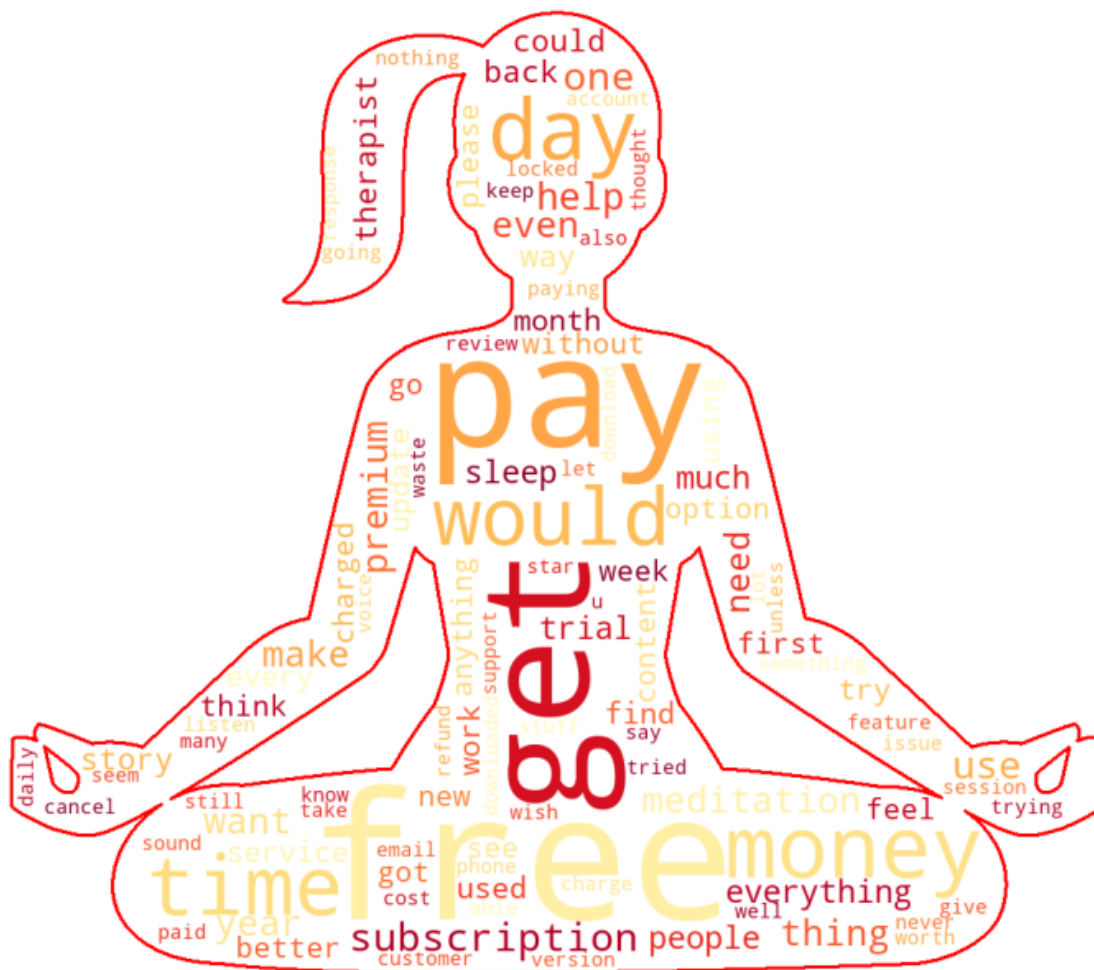
5.5.3 Binary Sentiment WordClouds

Let's take a look at whether our results differ when we split our data into two classes as opposed to three.

```
[66]: # Generate positive wordcloud for binary sentiment
bin_pos_tokens = wordcloud_prep(bin_pos_df['combined_text'], stopwords_list)
bin_pos_wc = generate_wordcloud(bin_pos_tokens, stopwords_list,
                                colormap='summer', background_color='white',
                                contour_color='green')
```



```
[67]: # Generate negative wordcloud for binary sentiment
bin_neg_tokens = wordcloud_prep(bin_neg_df['combined_text'], stopwords_list)
bin_neg_wc = generate_wordcloud(bin_neg_tokens, stopwords_list,
                                colormap='YlOrRd', background_color='white',
                                contour_color='red')
```



Again, we can see that our negative wordcloud contains lots of payment and customer-support related topics, but our positive wordcloud doesn't as clearly tell us what app features users appreciated. Not to worry though, since our modeling and interpretation should give us more accurate insight into what is considered positive or negative.

6 MODEL

During this phase, we will explore some different modeling algorithms to help us extract more accurate insights into what words are more influential in classifying sentiments.

For multi-class sentiments, we will explore Logistic Regression and Random Forest models. For binary sentiments, we will add Support Vector Classification, and where necessary, we will implement a gridsearch to optimize model performance and address the issue of under or overfitting to the training data.

As we saw during the Exploratory Data Analysis, there is a class imbalance for both the multi-class and binary data, and so we will set the class weight of all models to "balanced".

Since we saw that bigrams can reveal additional insight, we will also set our `ngram_range` to (1,2) to include vectorization of individual words as well as bigrams.

```
[68]: # Model evaluation function
def evaluate_model(y_test, y_train, X_test, X_train, clf, n_class=3,
                  figsize=(10,5), digits=2, params=False):
    """
    Displays evaluation metrics including classification report, confusion
    matrix, ROC-AUC curve.

    If the argument 'params' is passed, will display a table of the
    parameters hyperparameters used in the model.

    Args:
        y_test (Series) : Testing data with target variable.
        y_train (Series) : Training data with target variable.
        X_test (DataFrame) : Testing data with vectorized feature variables.
        X_train (DataFrame) : Training data with vectorized feature variables.
        clf (classifier object) : Type of classification model to use.
        n_class (int) : Number of classes for classification task.
        figsize (int, int) : Figure dimensions. Default is (10,5)
        digits (int) : Number of decimal places to display in classification
            report
        params (bool) : Prints table of hyperparameters used in model.

    Returns:
        """

    sns.set_style('dark')

    # Get Predictions
    y_hat_test = clf.predict(X_test)

    # Display training and test accuracy
    print("****TRAIN VS TEST ACCURACY****")
    print("\n", f"Training Score: {round(clf.score(X_train, y_train),2)}",
          "\n", f"Test Score: {round(clf.score(X_test, y_test),2)}", "\n")

    # Display classification report
    print("****CLASSIFICATION REPORT - TEST DATA****")
    print(metrics.classification_report(y_test, y_hat_test, digits=digits))
    print("****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****")

    if n_class == 3:
        # Create a figure/axes for confusion matrix and ROC curve
        fig, ax = plt.subplots(ncols=2, figsize=figsize)
```

```

# Plot confusion matrix
plot_confusion_matrix(estimator=clf, X=X_test, y_true=y_test,
                      cmap='Purples', normalize='true', ax=ax[0],
                      display_labels=['Negative', 'Neutral',
                                     'Positive'])

ax[0].set_title('Confusion Matrix')

# Plot ROC-AUC curve
pred_prob = clf.predict_proba(X_test)
fpr={}
tpr={}
thresh={}

for i in range(n_class):
    fpr[i], tpr[i], thresh[i] = roc_curve(y_test, pred_prob[:,i],
                                         pos_label=i)

ax = ax[1]
ax.grid()
ax.plot(fpr[0], tpr[0], linestyle='-',color='red', label='Negative')
ax.plot(fpr[1], tpr[1], linestyle='-',color='blue', label='Neutral')
ax.plot(fpr[2], tpr[2], linestyle='-',color='green', label='Positive')
ax.set_title('ROC-AUC curve')
ax.set_xlabel('False Positive Rate')
ax.set_ylabel('True Positive rate')
ax.legend(loc='best')
ax.plot([0,1], [0,1], ls='-', color='purple')

plt.tight_layout()
plt.show()

elif n_class == 2:
    fig, axes = plt.subplots(ncols=2,
                            figsize=figsize)

    # Plot confusion matrix
    metrics.plot_confusion_matrix(clf, X_test,
                                y_test,normalize='true',
                                cmap='Purples',ax=axes[0],
                                display_labels=['Negative', 'Positive'])
    axes[0].set_title('Confusion Matrix')

    # Plot ROC-AUC curve
    metrics.plot_roc_curve(clf,X_test,y_test,ax=axes[1])

    ax = axes[1]

```

```

ax.plot([0,1],[0,1], ls='-', color='purple')
ax.grid()
ax.set_title('ROC-AUC Curve')
ax.set_xlabel('False Positive Rate')
ax.set_ylabel('True Positive rate')
ax.get_legend().remove()

plt.tight_layout()
plt.show()

# Print model parameters
if params == True:
    print("****MODEL PARAMETERS****")
    params = pd.DataFrame(pd.Series(clf.get_params()))
    params.columns=['parameters']
    display(params)

```

6.1 Multi-Class Modeling: [1-2] Negative, [3] Neutral, [4-5] Positive

6.1.1 Vectorizing Train-Test Splits

We begin the modeling process by creating an X and y train-test split and then using a TF-IDF vectorizer to preprocess each document.

```

[69]: # Make X and y
y_multi = clean_df['multi_sent'].copy()
X = clean_df['combined_text'].copy()

# Train Test Split
X_multi_train, X_multi_test, y_multi_train, y_multi_test = train_test_split(X,
    ↪y_multi,
    ↪test_size=0.3,
    ↪random_state=27)

# Check y_train value counts
y_multi_train.value_counts(normalize=True)

```

```

[69]: 2    0.774930
      0    0.171951
      1    0.053119
      Name: multi_sent, dtype: float64

```

```

[70]: # Initialize RegExp tokenizer
tokenizer = RegexpTokenizer(r"([a-zA-Z]+(?:'[a-z]+)?)")

```

```
# Make a TF-IDF vectorizer
vectorizer_multi = TfidfVectorizer(tokenizer=tokenizer.tokenize,
                                   token_pattern=r"([a-zA-Z]+(?:'[a-z]+)?)",
                                   stop_words=stopwords_list,
                                   ngram_range=(1,2))

# Vectorize data and make X_train_tfidf and X_test_tfidf
X_multi_train_tfidf = vectorizer_multi.fit_transform(X_multi_train)
X_multi_test_tfidf = vectorizer_multi.transform(X_multi_test)
X_multi_train_tfidf
```

```
[70]: <31288x310249 sparse matrix of type '<class 'numpy.float64'>'
      with 1030078 stored elements in Compressed Sparse Row format>
```

6.1.2 Logistic Regression

We are now ready to fit our first model using a Logistic Regression algorithm on our multi-class data. Due to hardware limitations as well as the fact that we will later see better performing models, we will omit the gridsearch phase for this particular model.

```
[71]: # Fit Logistic Regression model on multi-class training data
logreg = LogisticRegression(max_iter=200, class_weight='balanced')
logreg.fit(X_multi_train_tfidf, y_multi_train)
```

```
[71]: LogisticRegression(class_weight='balanced', max_iter=200)
```

```
[72]: # Evaluate model performance
evaluate_model(y_multi_test, y_multi_train,
              X_multi_test_tfidf, X_multi_train_tfidf, logreg)
```

****TRAIN VS TEST ACCURACY****

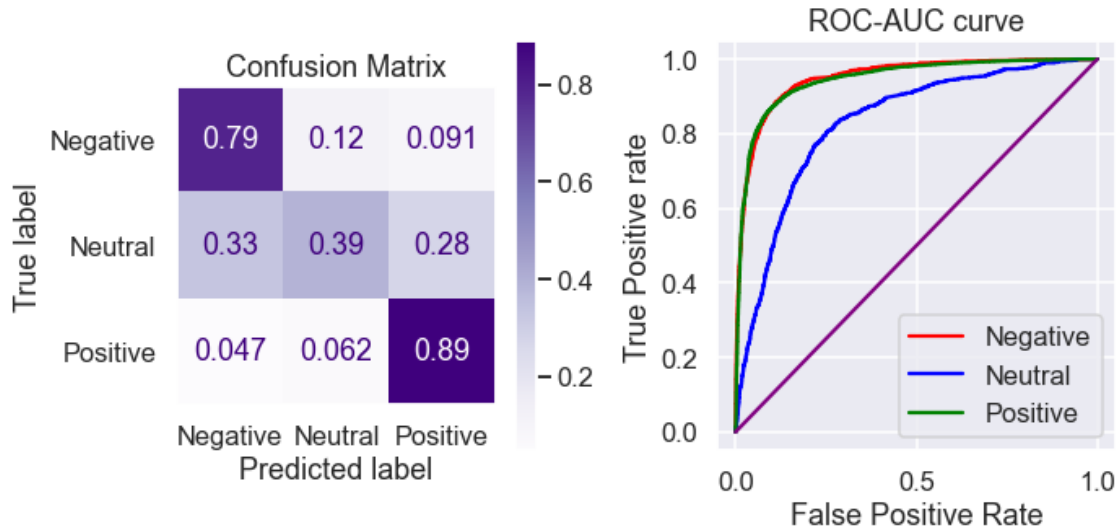
Training Score: 0.95

Test Score: 0.85

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.72	0.79	0.75	2347
1	0.23	0.39	0.29	717
2	0.96	0.89	0.92	10346
accuracy			0.85	13410
macro avg	0.64	0.69	0.66	13410
weighted avg	0.88	0.85	0.86	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



Even without the help of a gridsearch, our Logistic Regression model already seems to have decent performance when it comes to classifying negative and positive sentiments. It does have much poorer performance classifying neutral sentiments, but this is understandable considering that the neutral sentiment wordcloud was also not able to give us much insight into what topics were associated with a neutral sentiment.

6.1.3 Random Forest

```
[73]: # Fit Random Forest model on multi-class training data
rf = RandomForestClassifier(class_weight='balanced')
rf.fit(X_multi_train_tfidf, y_multi_train)
```

```
[73]: RandomForestClassifier(class_weight='balanced')
```

```
[74]: # Evaluate model performance
evaluate_model(y_multi_test, y_multi_train, X_multi_test_tfidf,
               X_multi_train_tfidf, rf)
```

****TRAIN VS TEST ACCURACY****

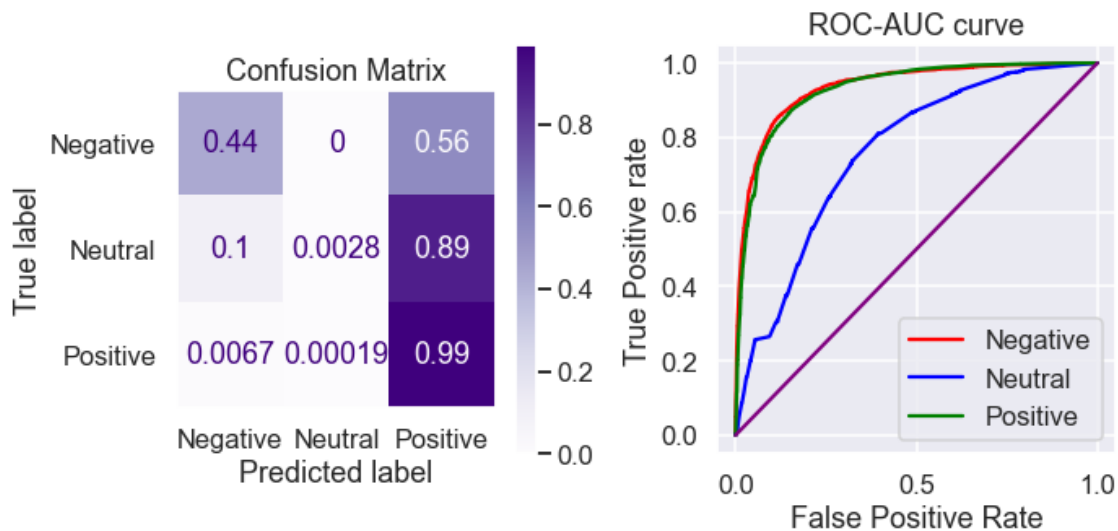
Training Score: 1.0
Test Score: 0.84

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.88	0.44	0.59	2347
1	0.50	0.00	0.01	717
2	0.84	0.99	0.91	10346

accuracy			0.84	13410
macro avg	0.74	0.48	0.50	13410
weighted avg	0.83	0.84	0.81	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



On our first iteration of a Random Forest model, we can see that the model is clearly being overfit to the training data, resulting in extremely poor recall scores for negative and neutral sentiments. Let's use a gridsearch to address this issue.

```
[75]: # Initialize Random Forest algorithm
rf = RandomForestClassifier(class_weight='balanced')

# Create hyperparameter grid to gridsearch
param_grid = {'criterion': ['gini', 'entropy'],
              'max_depth': [10, 20, 40],
              'n_estimators': [50, 100, 200],
              'min_samples_leaf': [2, 5, 10]}

# Initialize gridsearch optimized for F1 macro score
gridsearch = GridSearchCV(estimator=rf, param_grid = param_grid,
                          scoring='f1_macro', n_jobs=-1)

# Run gridsearch and display best hyperparameters
gridsearch.fit(X_multi_train_tfidf, y_multi_train)
gridsearch.best_params_
```

```
[75]: {'criterion': 'entropy',
      'max_depth': 40,
```

```
'min_samples_leaf': 2,
'n_estimators': 200}
```

```
[76]: # Evaluate best model performance
evaluate_model(y_multi_test, y_multi_train, X_multi_test_tfidf,
               X_multi_train_tfidf, gridsearch.best_estimator_)
```

****TRAIN VS TEST ACCURACY****

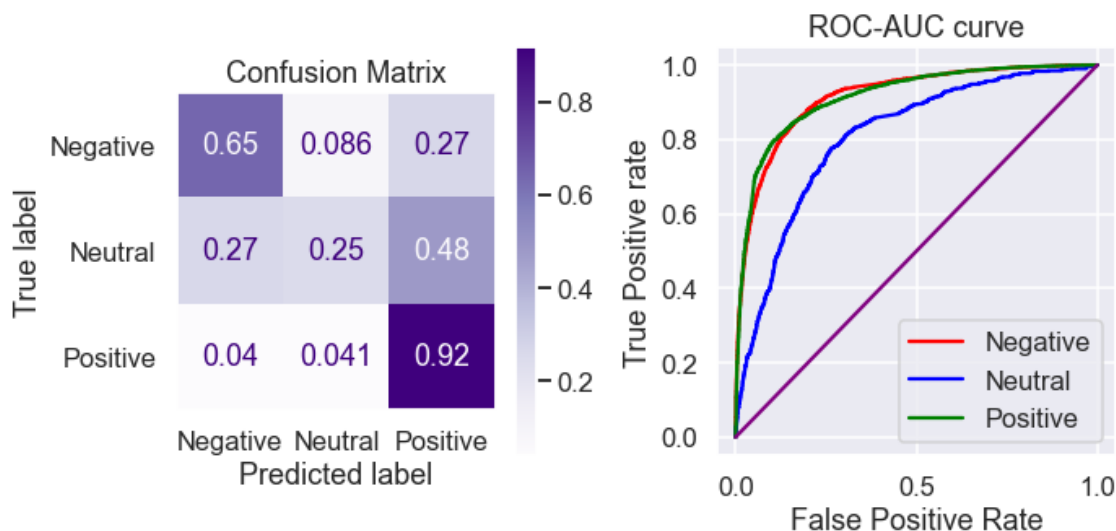
Training Score: 0.87

Test Score: 0.84

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.72	0.65	0.68	2347
1	0.22	0.25	0.24	717
2	0.91	0.92	0.91	10346
accuracy			0.84	13410
macro avg	0.62	0.61	0.61	13410
weighted avg	0.84	0.84	0.84	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



Our gridsearched model's performance has improved considerably in classifying negative and neutral sentiments, but it still isn't quite ideal. Let's see if running another gridsearch can further improve performance.

```
[77]: # Initialize Random Forest algorithm
rf = RandomForestClassifier(class_weight='balanced')

# Create hyperparameter grid to gridsearch
param_grid = {'criterion': ['entropy'],
              'max_depth': [40, 60],
              'n_estimators': [200, 300],
              'min_samples_leaf': [5]}

# Initialize gridsearch optimized for F1 macro score
gridsearch = GridSearchCV(estimator=rf, param_grid = param_grid,
                          scoring='f1_macro', n_jobs=-1)

# Run gridsearch and display best hyperparameters
gridsearch.fit(X_multi_train_tfidf, y_multi_train)
gridsearch.best_params_
```

```
[77]: {'criterion': 'entropy',
      'max_depth': 60,
      'min_samples_leaf': 5,
      'n_estimators': 300}
```

```
[78]: # Evaluate best model performance
evaluate_model(y_multi_test, y_multi_train,
              X_multi_test_tfidf, X_multi_train_tfidf,
              gridsearch.best_estimator_)
```

****TRAIN VS TEST ACCURACY****

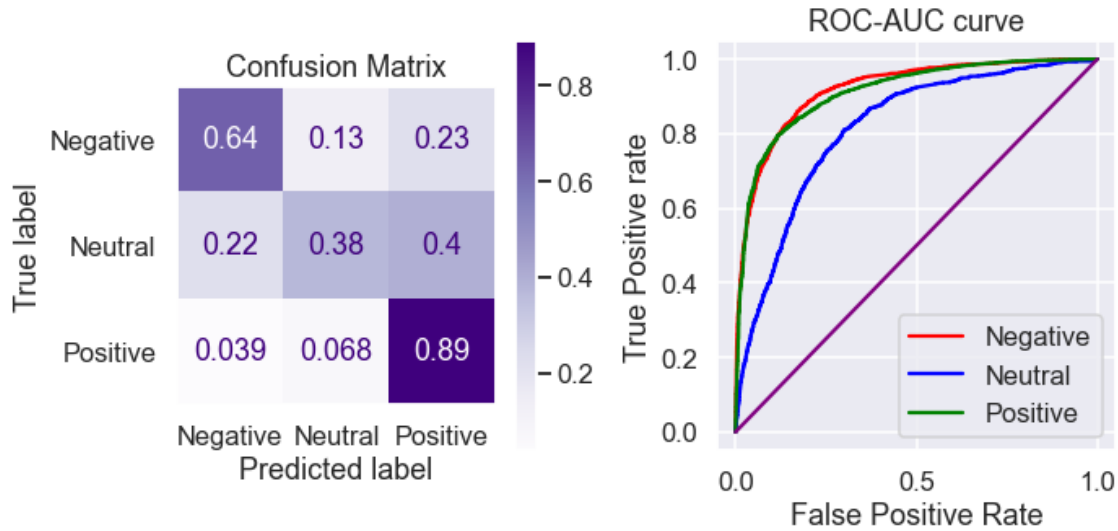
Training Score: 0.84

Test Score: 0.82

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.73	0.64	0.68	2347
1	0.21	0.38	0.27	717
2	0.92	0.89	0.91	10346
accuracy			0.82	13410
macro avg	0.62	0.64	0.62	13410
weighted avg	0.85	0.82	0.83	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



We do see some improvement in the model's ability to correctly predict neutral sentiment, but this comes at the cost of our negative and positive sentiment recall scores. Hence, we conclude that the Logistic Regression algorithm is a much better choice than the Random Forest algorithm for our task.

6.2 Binary Modeling: [1-3] Negative, [4-5] Positive

Here, we return to our binary classification problem where ratings of 1-3 stars will be considered to be negative, and 4-5 star ratings will be considered to be positive.

6.2.1 Vectorizing Train-Test Splits

Similar to the multi-class modeling problem, we will vectorize the data following its new train-test split.

```
[79]: # Make y
y_bin = clean_df['sent_bin_4up'].copy()

## Train Test Split
X_bin_train, X_bin_test, y_bin_train, y_bin_test = train_test_split(X, y_bin,
                                                                    test_size=0.
                                                                    ↪3,

                                                                    ↪random_state=27)

# Check y_train value counts
y_bin_train.value_counts(normalize=True)
```

```
[79]: 1    0.77493
      0    0.22507
```

Name: sent_bin_4up, dtype: float64

```
[80]: # Make a TF-IDF vectorizer using previously created tokenizer
vectorizer_bin = TfidfVectorizer(tokenizer=tokenizer.tokenize,
                                token_pattern=r"([a-zA-Z]+(?:'[a-z]+)?)",
                                stop_words=stopwords_list,
                                ngram_range=(1,2))

# Vectorize data and make X_train_tfidf and X_test_tfidf
X_bin_train_tfidf = vectorizer_bin.fit_transform(X_bin_train)
X_bin_test_tfidf = vectorizer_bin.transform(X_bin_test)
X_bin_train_tfidf
```

```
[80]: <31288x310249 sparse matrix of type '<class 'numpy.float64'>'
      with 1030078 stored elements in Compressed Sparse Row format>
```

6.2.2 Logistic Regression

Again, let's start modeling for our binary classification problem with Logistic Regression.

```
[81]: # Fit Logistic Regression model on binary training data
logreg = LogisticRegression(class_weight='balanced')
logreg.fit(X_bin_train_tfidf, y_bin_train)
```

```
[81]: LogisticRegression(class_weight='balanced')
```

```
[82]: # Evaluate model performance
evaluate_model(y_bin_test, y_bin_train,
              X_bin_test_tfidf, X_bin_train_tfidf, logreg, 2)
```

****TRAIN VS TEST ACCURACY****

Training Score: 0.94

Test Score: 0.89

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.73	0.86	0.79	3064
1	0.96	0.90	0.93	10346
accuracy			0.89	13410
macro avg	0.84	0.88	0.86	13410
weighted avg	0.90	0.89	0.90	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



Right off the bat, we see great performance with our Logistic Regression algorithm. Again, due to hardware limitations we will not gridsearch this model. Although the model does seem to be slightly overfitting to the training data, it is not overfitting by a large amount. We can see that our F1 score is 0.86 for negative classification and .90 for positive classification, which is quite good.

Let's try a cross validated iteration of the Logistic Regression to see if we can improve our results on the test data.

```
[83]: # Fit Logistic Regression model on binary training data
logregcv = LogisticRegressionCV(max_iter=750, class_weight='balanced')
logregcv.fit(X_bin_train_tfidf, y_bin_train)
```

```
[83]: LogisticRegressionCV(class_weight='balanced', max_iter=750)
```

```
[84]: # Evaluate model performance
evaluate_model(y_bin_test, y_bin_train,
               X_bin_test_tfidf, X_bin_train_tfidf, logregcv, 2)
```

****TRAIN VS TEST ACCURACY****

Training Score: 1.0

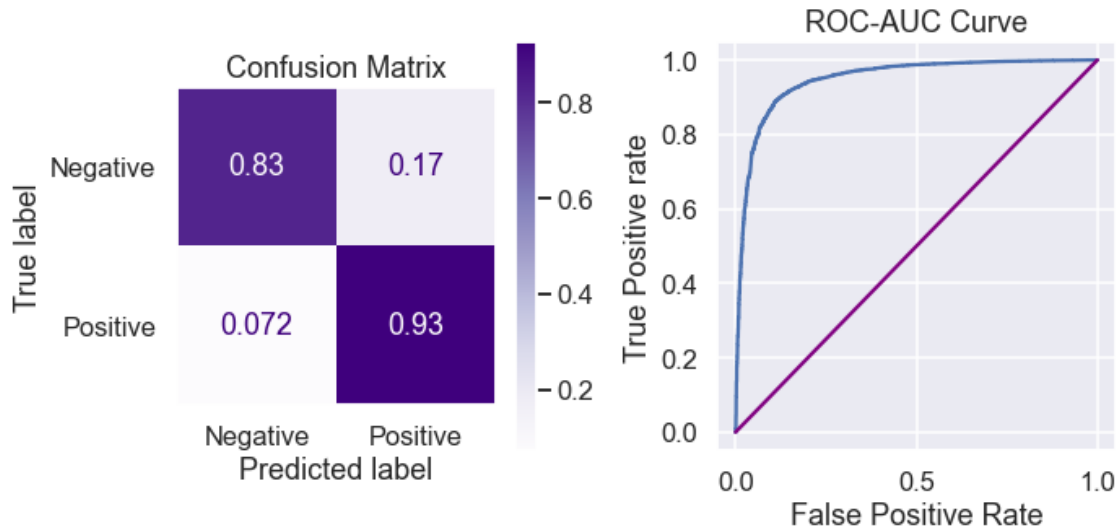
Test Score: 0.9

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.77	0.83	0.80	3064
1	0.95	0.93	0.94	10346
accuracy			0.90	13410

macro avg	0.86	0.88	0.87	13410
weighted avg	0.91	0.90	0.91	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



We see that although this model has higher predictive capability on positive sentiment, it comes at the cost of a lower recall score for negative sentiment, as well as a more overfit model than we saw before. In order to preserve a more even prediction rate between negative and positive sentiments, we will proceed with the non-cross-validated version of Logistic Regression as our best model.

6.2.3 Random Forest

```
[85]: # Fit Random Forest model on binary training data
      rf = RandomForestClassifier(class_weight='balanced')
      rf.fit(X_bin_train_tfidf, y_bin_train)
```

```
[85]: RandomForestClassifier(class_weight='balanced')
```

```
[86]: # Evaluate model performance
      evaluate_model(y_bin_test, y_bin_train,
                    X_bin_test_tfidf, X_bin_train_tfidf, rf, 2)
```

****TRAIN VS TEST ACCURACY****

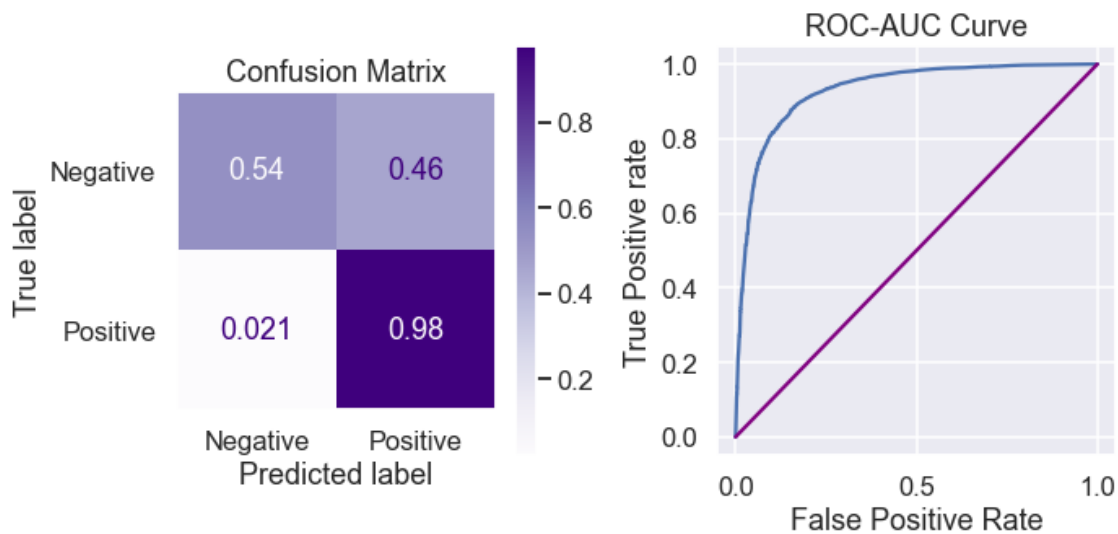
Training Score: 1.0
Test Score: 0.88

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.88	0.54	0.67	3064
1	0.88	0.98	0.93	10346
accuracy			0.88	13410
macro avg	0.88	0.76	0.80	13410
weighted avg	0.88	0.88	0.87	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



Because we haven't altered any of the hyperparameters for our Random Forest algorithm, we again see that our modelling is clearly overfitting to the training data. Let's run a few gridsearches on this model.

```
[87]: # Initialize Random Forest algorithm
rf = RandomForestClassifier(class_weight='balanced')

# Create hyperparameter grid to gridsearch
param_grid = {'criterion': ['gini', 'entropy'],
              'max_depth': [10, 20, 30, 40],
              'min_samples_leaf': [3, 5, 7]}

# Initialize gridsearch optimized for F1 macro score
gridsearch = GridSearchCV(estimator=rf, param_grid = param_grid,
                          scoring='f1_macro', n_jobs=-1)

# Run gridsearch and display best hyperparameters
gridsearch.fit(X_bin_train_tfidf, y_bin_train)
gridsearch.best_params_
```

```
[87]: {'criterion': 'entropy', 'max_depth': 40, 'min_samples_leaf': 3}
```

```
[88]: # Evaluate model performance
evaluate_model(y_bin_test, y_bin_train,
               X_bin_test_tfidf, X_bin_train_tfidf,
               gridsearch.best_estimator_, 2)
```

****TRAIN VS TEST ACCURACY****

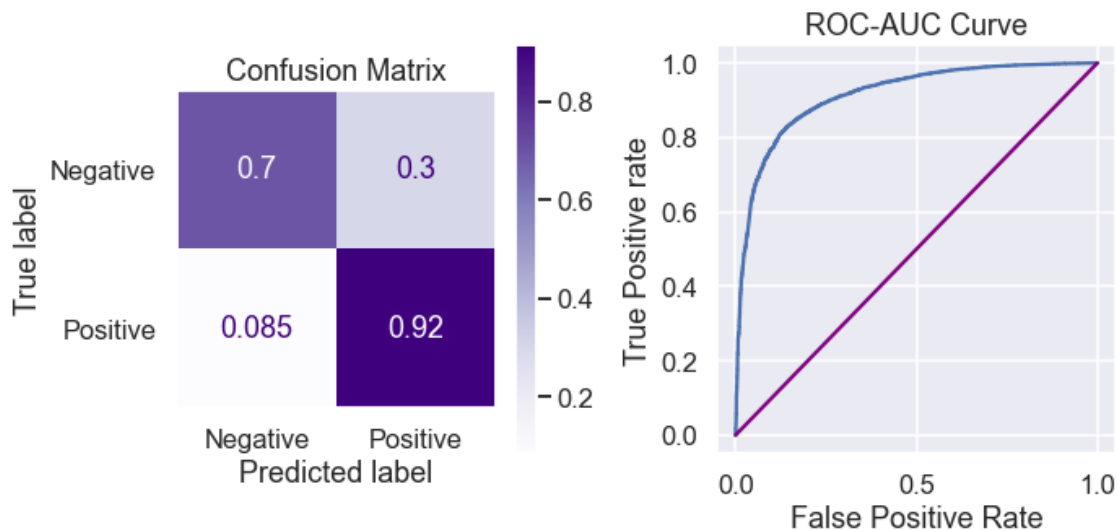
Training Score: 0.87

Test Score: 0.87

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.71	0.70	0.70	3064
1	0.91	0.92	0.91	10346
accuracy			0.87	13410
macro avg	0.81	0.81	0.81	13410
weighted avg	0.86	0.87	0.86	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



Great! We have addressed the issue of overfitting. Our model is performing much better on the negative classification, and only a little worse on the positive classification. Let's see if we can further improve our gridsearch performance.

```
[89]: # Initialize Random Forest algorithm
rf = RandomForestClassifier(class_weight='balanced')
```

```

# Create hyperparameter grid to gridsearch
param_grid = {'criterion': ['entropy'],
              'max_depth': [40, 50, 60],
              'min_samples_leaf': [1, 2, 3]}

# Initialize gridsearch optimized for F1 macro score
gridsearch = GridSearchCV(estimator=rf, param_grid = param_grid,
                          scoring='f1_macro', n_jobs=-1)

# Run gridsearch and display best hyperparameters
gridsearch.fit(X_bin_train_tfidf, y_bin_train)
gridsearch.best_params_

```

```
[89]: {'criterion': 'entropy', 'max_depth': 60, 'min_samples_leaf': 1}
```

```

[90]: # Evaluate model performance
evaluate_model(y_bin_test, y_bin_train,
              X_bin_test_tfidf, X_bin_train_tfidf,
              gridsearch.best_estimator_, 2)

```

****TRAIN VS TEST ACCURACY****

Training Score: 0.95

Test Score: 0.88

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.80	0.64	0.71	3064
1	0.90	0.95	0.93	10346
accuracy			0.88	13410
macro avg	0.85	0.80	0.82	13410
weighted avg	0.88	0.88	0.88	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



Again, we can see that we are already overfitting to the training data again, and the model's overall performance is already getting worse.

6.2.4 Support Vector Classification

Because our vectorized bag of words has a much higher number of columns than rows, we can expect our Support Vector Classification model to perform well.

```
[91]: # Fit SVC model on binary training data
svc = LinearSVC()
svc.fit(X_bin_train_tfidf, y_bin_train)
```

```
[91]: LinearSVC()
```

```
[92]: # Evaluate model performance
evaluate_model(y_bin_test, y_bin_train,
               X_bin_test_tfidf, X_bin_train_tfidf, svc, 2)
```

****TRAIN VS TEST ACCURACY****

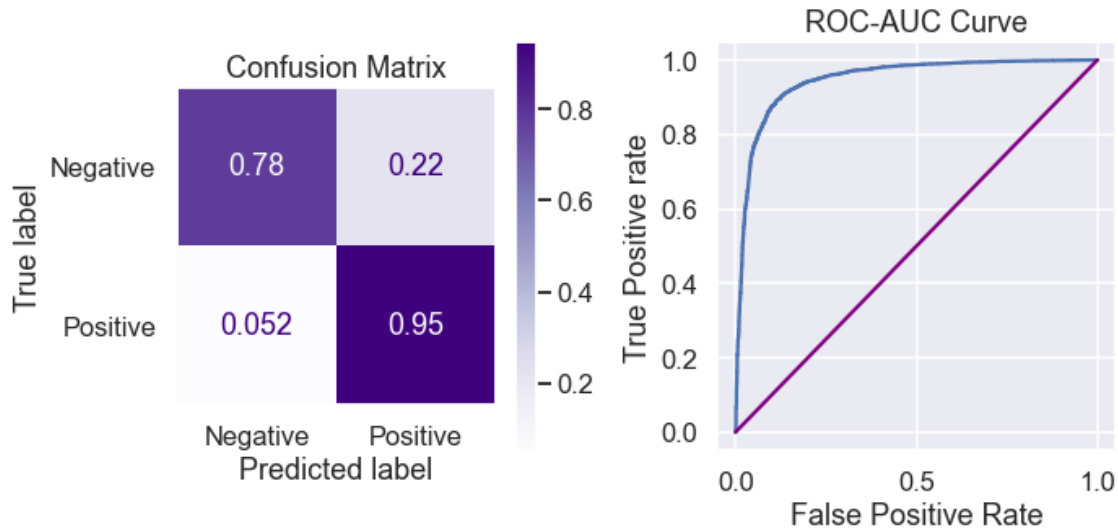
```
Training Score: 1.0
Test Score: 0.91
```

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.81	0.78	0.80	3064
1	0.93	0.95	0.94	10346
accuracy			0.91	13410

macro avg	0.87	0.86	0.87	13410
weighted avg	0.91	0.91	0.91	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



Without altering the model's hyperparameters, we can already see decent performance, with positive classification performing slightly better than our Logistic Regression, and negative classification doing slightly worse. Let's use a gridsearch to see if we can improve the model's performance.

```
[93]: # Initialize SVC algorithm
svc = LinearSVC(class_weight='balanced')

# Create hyperparameter grid to gridsearch
param_grid = {'penalty': ['l1', 'l2'],
              'loss': ['hinge', 'squared_hinge'],
              'tol': [0.000001, 0.0001, 0.01],
              'C': [0.1, 1, 10],
              'fit_intercept': [True, False]}

# Initialize gridsearch optimized for F1 macro score
gridsearch = GridSearchCV(estimator=svc, param_grid = param_grid,
                          scoring='f1_macro', n_jobs=-1)

# Run gridsearch and display best hyperparameters
gridsearch.fit(X_bin_train_tfidf, y_bin_train)
gridsearch.best_params_
```

```
[93]: {'C': 1, 'fit_intercept': True, 'loss': 'hinge', 'penalty': 'l2', 'tol': 1e-06}
```

```
[94]: # Evaluate best model's performance
evaluate_model(y_bin_test, y_bin_train,
               X_bin_test_tfidf, X_bin_train_tfidf,
               gridsearch.best_estimator_, 2)
```

****TRAIN VS TEST ACCURACY****

Training Score: 0.98

Test Score: 0.9

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.76	0.85	0.80	3064
1	0.95	0.92	0.94	10346
accuracy			0.90	13410
macro avg	0.85	0.88	0.87	13410
weighted avg	0.91	0.90	0.90	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



7 iNTERPRET

Now that we're done modeling, we can move on to interpreting the results. Below, we can see a summary of our best performing models, which were the Logistic Regression for multi-class sentiments, and Logistic Regression for binary sentiments. Out of the two models, our multi-class model had poorer performance, and so we will only be using that as reference to support any findings from the best model. Although our SVC model did have a slightly better F1 macro score,

we will focus on the Logistic Regression model due to its more even predictive capability between negative and positive sentiments and also for ease of interpretation.

7.0.1 Best Model Summary

By observing the evaluation metrics from each of our models, we can see that the binary classification models had much better performance. We will still examine the feature coefficients found in our Multi-Class Logistic Regression as a reference point, but for extracting final insights, we will make use of our Binary Logistic Regression and SVC models. Below, we can find a roundup of each model's performance metrics along with their fit times.

Multi-Class Logistic Regression

```
[95]: # Fit best performing Logistic Regression on multi-class data
# and display performance
%time
best_multi_lr = LogisticRegression(max_iter=200, class_weight='balanced')
best_multi_lr.fit(X_multi_train_tfidf, y_multi_train)
evaluate_model(y_multi_test, y_multi_train,
               X_multi_test_tfidf, X_multi_train_tfidf, best_multi_lr)
```

CPU times: user 2 μ s, sys: 1 μ s, total: 3 μ s

Wall time: 5.01 μ s

****TRAIN VS TEST ACCURACY****

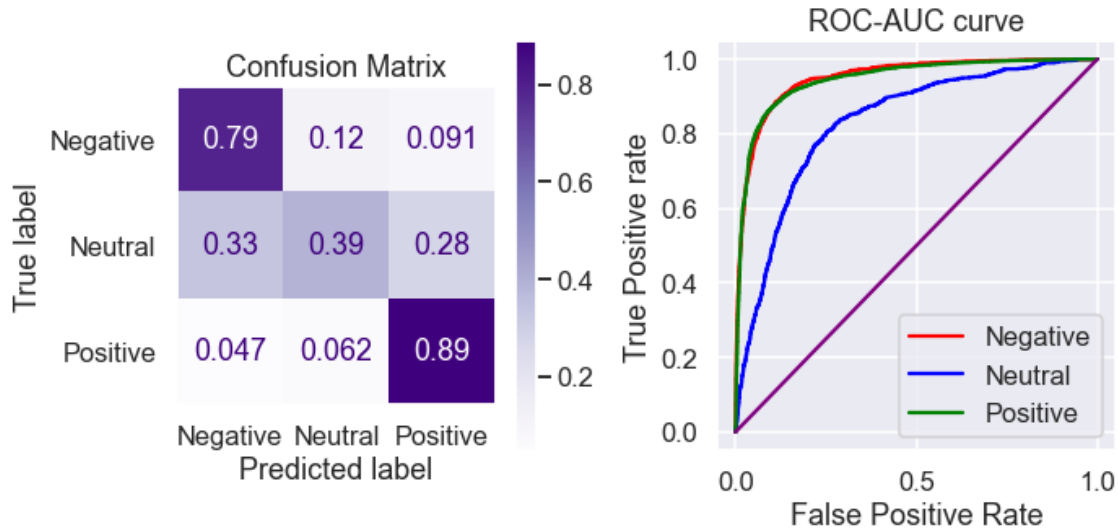
Training Score: 0.95

Test Score: 0.85

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.72	0.79	0.75	2347
1	0.23	0.39	0.29	717
2	0.96	0.89	0.92	10346
accuracy			0.85	13410
macro avg	0.64	0.69	0.66	13410
weighted avg	0.88	0.85	0.86	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



Binary Logistic Regression

```
[96]: # Fit best performing Logistic Regression on binary data
# and display performance
%time
best_bin_lr = LogisticRegression(class_weight='balanced')
best_bin_lr.fit(X_bin_train_tfidf, y_bin_train)
evaluate_model(y_bin_test, y_bin_train,
               X_bin_test_tfidf, X_bin_train_tfidf, best_bin_lr, 2)
```

CPU times: user 2 μ s, sys: 0 ns, total: 2 μ s

Wall time: 5.01 μ s

****TRAIN VS TEST ACCURACY****

Training Score: 0.94

Test Score: 0.89

****CLASSIFICATION REPORT - TEST DATA****

	precision	recall	f1-score	support
0	0.73	0.86	0.79	3064
1	0.96	0.90	0.93	10346
accuracy			0.89	13410
macro avg	0.84	0.88	0.86	13410
weighted avg	0.90	0.89	0.90	13410

****CONFUSION MATRIX AND ROC-AUC VISUALIZATION****



We can see that there is a small difference in training times between these models.

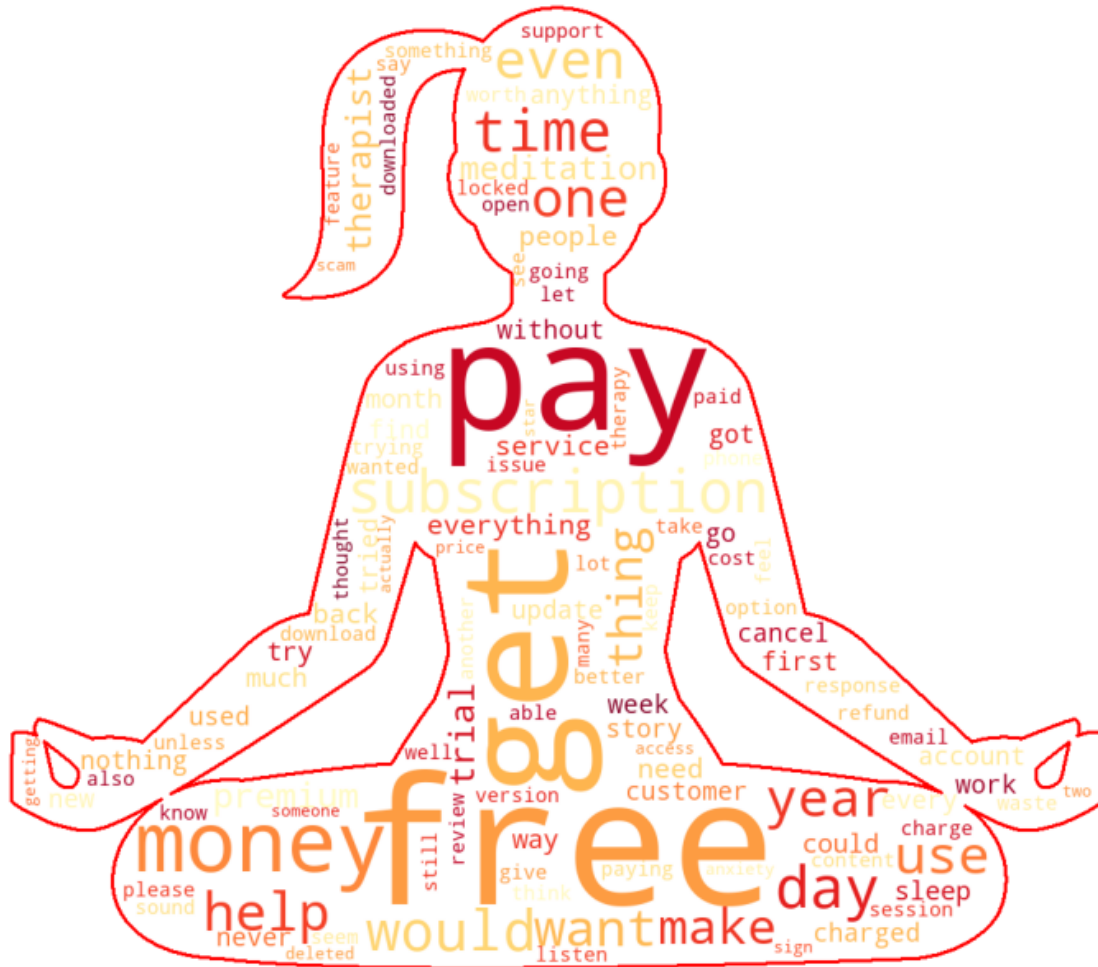
7.0.2 Evaluating Feature Importances to Sentiment Classification

Here, we provide a summary of visualizations that can help us determine what words are indicative of either positive or negative sentiments. We first revisit our wordclouds to remind ourselves what words are frequently found in negative and positive sentiments, then move on to examine feature coefficients from our Logistic Regression models.

Revisiting Multi-Class WordClouds

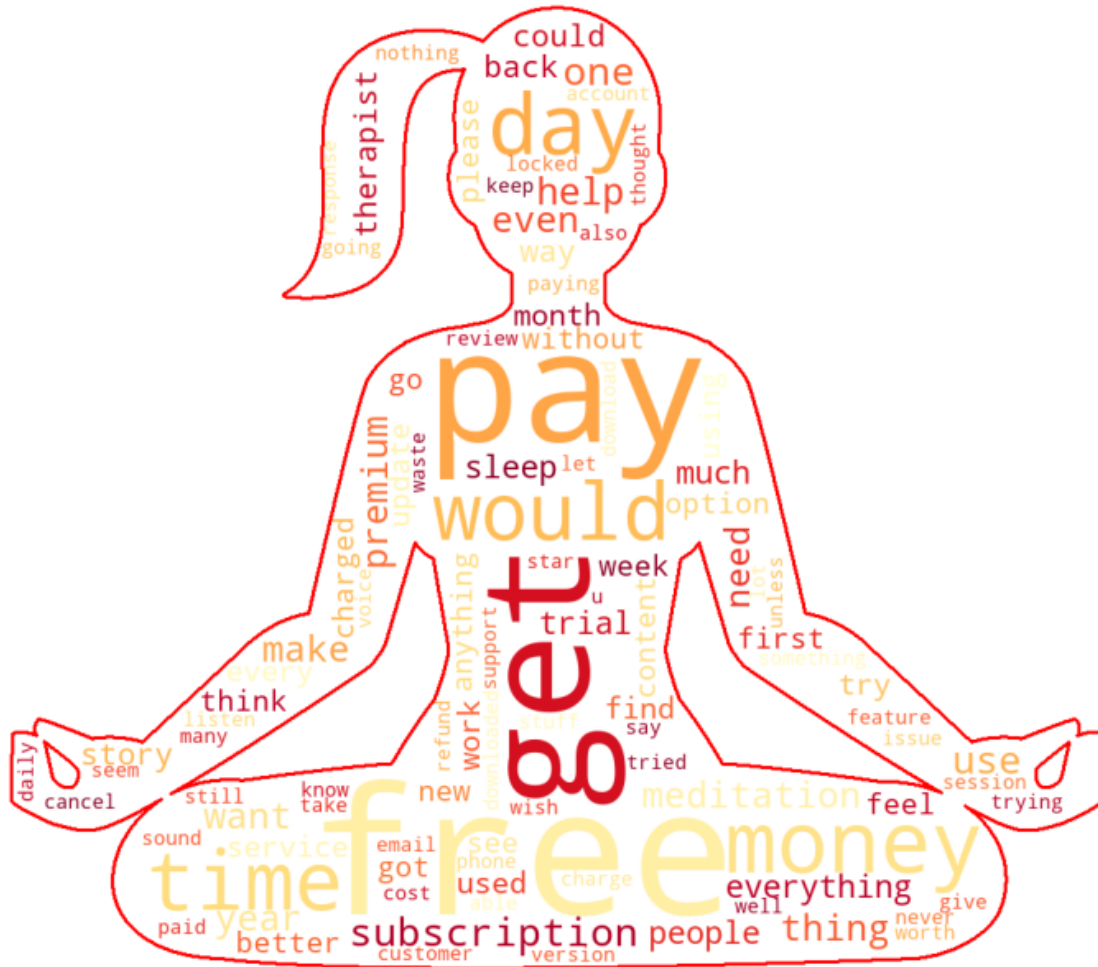
```
[97]: plt.figure(figsize = (12, 12), facecolor = None)
plt.imshow(multi_neg_wc)
plt.title('Multi-Class Negative WordCloud')
plt.axis('off');
```

Multi-Class Negative WordCloud



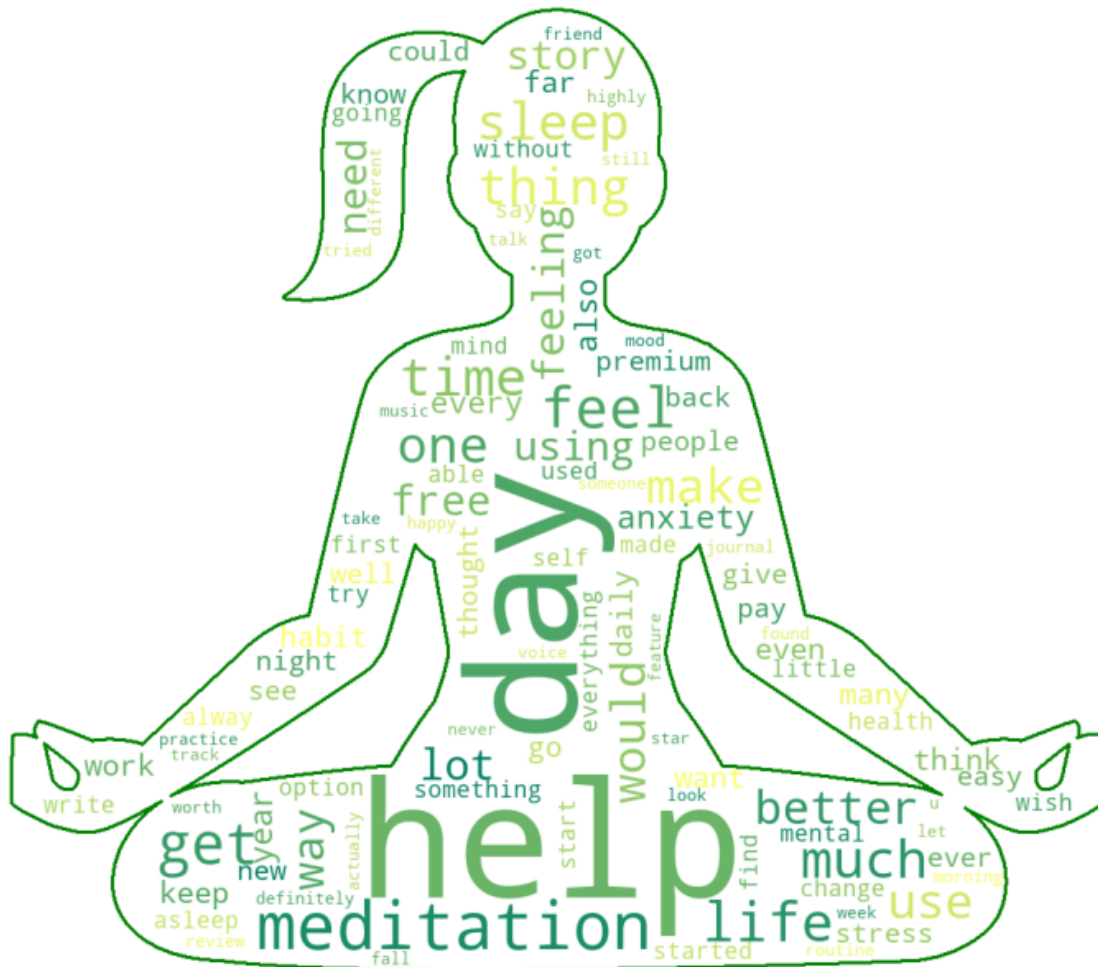
```
[98]: plt.figure(figsize = (12, 12), facecolor = None)
plt.imshow(bin_neg_wc)
plt.title('Binary Negative WordCloud')
plt.axis('off');
```

Binary Negative WordCloud



```
[99]: plt.figure(figsize = (12, 12), facecolor = None)
plt.imshow(multi_pos_wc)
plt.title('Multi-Class Positive WordCloud')
plt.axis('off');
```

Multi-Class Positive WordCloud



```
[100]: plt.figure(figsize = (12, 12), facecolor = None)
plt.imshow(bin_pos_wc)
plt.title('Binary Positive WordCloud')
plt.axis('off');
```

In order to easily visualize what words had a heavier impact on the classification tasks, we will create barplots of the feature coefficients. For Logistic Regression, our units are in log odds, and so we can use the resulting visuals to get an idea of how much importance certain words relative to other words.

```
[101]: # Extract coefficients for negative classification from multi-class LogReg model
feature_names_multi = vectorizer_multi.get_feature_names()
log_coeff_multi = pd.Series(best_multi_lr.coef_[0],
                           index=feature_names_multi).
    ↪ sort_values(ascending=False)
lr_multi_neg = log_coeff_multi.head(30)
```

```
lr_multi_neg
```

```
[101]: waste      4.149088
      pay       3.930401
      scam     3.911700
      refund   3.752736
      nothing  3.589443
      useless  3.360529
      money    3.357976
      cancel   3.322177
      charged  3.271018
      free     3.228345
      paid     3.165171
      youtube  3.060643
      deleted  3.060080
      even     2.959672
      anything 2.938652
      response 2.908503
      charge   2.851537
      service  2.808596
      trial    2.759657
      can't    2.556559
      used     2.437097
      unless   2.404335
      downloaded 2.376491
      subscription 2.364908
      waste time 2.360954
      another  2.343165
      account  2.250450
      support  2.167510
      ridiculous 2.138987
      tried    2.107287
      dtype: float64
```

```
[102]: # Extract coefficients for neutral classification from multi-class LogReg model
      log_coeff_multi = pd.Series(best_multi_lr.coef_[1],
                                   index=feature_names_multi).
      ↪sort_values(ascending=False)
      lr_multi_neut = log_coeff_multi.head(30)
      lr_multi_neut
```

```
[102]: wish      3.306832
      however  3.296086
      needs    3.156431
      okay     2.619824
      pay      2.542509
      locked   2.497298
```

update	2.487984
things	2.481749
it's ok	2.349625
doesn't	2.262670
notifications	2.238217
ok	2.211682
please	2.204475
premium	2.168191
three stars	2.163153
three	2.141338
meh	2.141167
could better	2.109335
though	2.084982
alright	2.053412
sometimes	2.012195
it's calming	2.008919
mostly	1.985532
think	1.960752
rest	1.955075
distracting	1.944241
ss	1.922132
joann	1.922132
fatima	1.905458
coaching	1.892594

dtype: float64

```
[103]: # Extract coefficients for positive classification from multi-class LogReg model
log_coeff_multi = pd.Series(best_multi_lr.coef_[2],
                             index=feature_names_multi).
↳ sort_values(ascending=False)
lr_multi_pos = log_coeff_multi.head(30)
lr_multi_pos
```

```
[103]: life          4.564987
easy            3.136781
highly         3.133140
day            2.680304
mind           2.511738
feel           2.323330
definitely     2.298280
far            2.192939
lot            2.190680
didn't pay    2.135592
ever           2.050639
truly          2.039540
grateful       2.026560
absolutely     1.929679
```

anxiety	1.907714
feelings	1.875360
friend	1.843545
meditation	1.818306
everyday	1.795276
i've	1.792564
practice	1.752999
feeling	1.727808
asleep	1.726594
happy	1.698626
gives	1.682562
night	1.649463
fun	1.619550
diary	1.610030
positive	1.570776
relaxed	1.561558

dtype: float64

```
[104]: # Extract coefficients for positive classification from binary LogReg model
feature_names_bin = vectorizer_bin.get_feature_names()
log_coeff_bin = pd.Series(best_bin_lr.coef_.flatten(),
                        index=feature_names_bin).sort_values(ascending=False)
lr_bin_pos = log_coeff_bin.head(22)
lr_bin_pos
```

```
[104]: life          6.675193
easy          4.320892
highly       4.144591
day          4.031501
feel         3.533113
far          3.494255
definitely   3.296762
mind         3.275490
lot          3.157402
feelings     2.989367
grateful     2.850748
truly        2.838214
meditation   2.760275
didn't pay   2.700208
i've         2.595948
feeling      2.559657
asleep       2.510027
everyday     2.506764
anxiety      2.504001
friend       2.476129
habits       2.455346
night        2.428035
```


dtype: float64

```
[105]: # Extract coefficients for negative classification from binary LogReg model
lr_bin_neg = log_coeff_bin.tail(22)
lr_bin_neg
```

```
[105]: refund          -4.147979
      paying          -4.158967
      useless         -4.185140
      charge          -4.238370
      service         -4.249729
      frustrating     -4.288553
      scam            -4.348352
      waste           -4.404109
      crashes         -4.414699
      trial           -4.461804
      cancel          -4.746034
      subscription    -4.789707
      nothing         -5.263785
      can't           -5.267615
      deleted         -5.293287
      doesn't         -5.322337
      update          -5.366532
      locked          -5.800127
      charged         -5.837720
      free            -6.670681
      money           -7.534205
      pay             -9.175482
dtype: float64
```

Multi-Class Logistic Regression Coefficients

```
[106]: # Create function to create barplots using feature coefficients
def barplot_series(series, title, xlabel, ylabel, palette='cool'):
    """
    Displays barplot of input series.

    If the argument 'params' is passed, will display a table of the
    parameters hyperparameters used in the model.

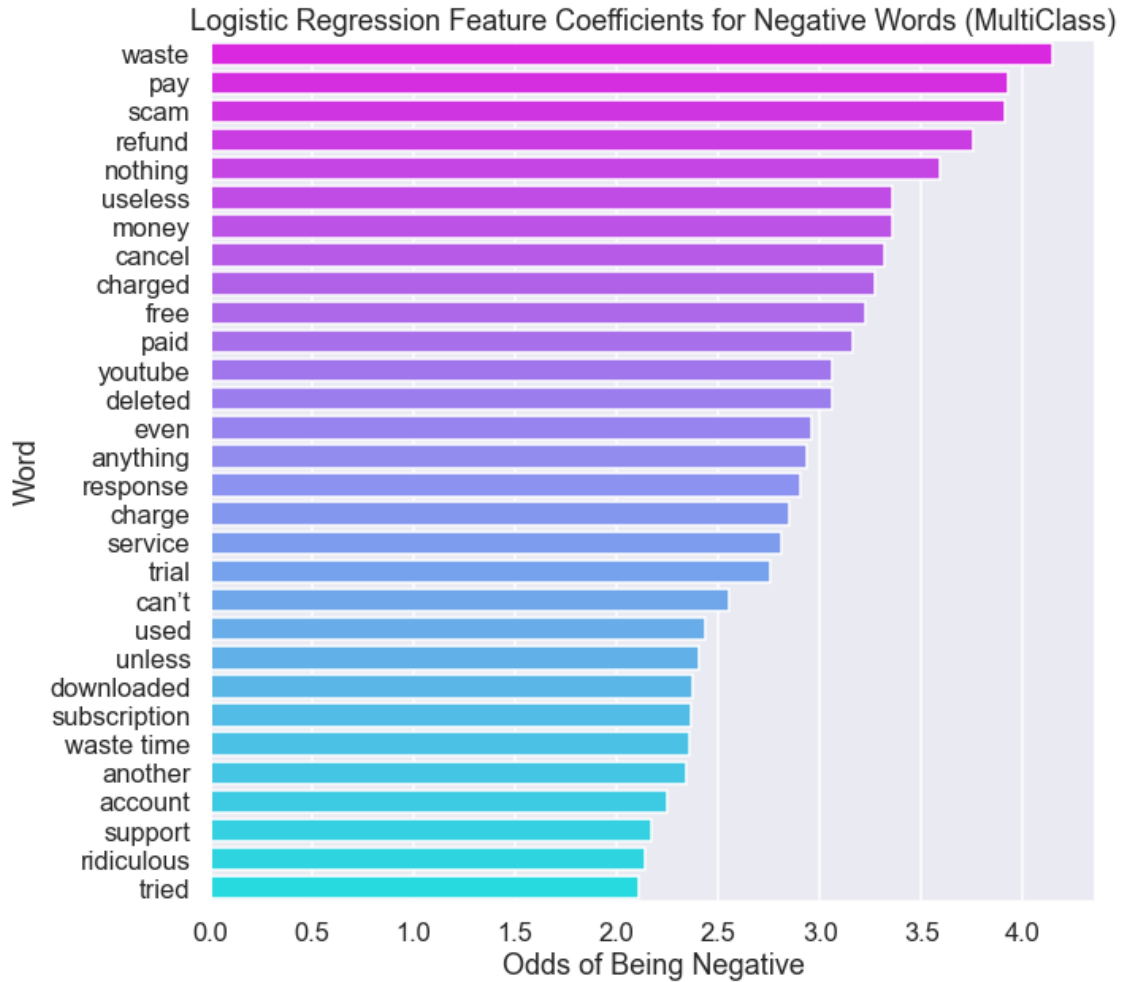
    Args:
        series (Series) : Series containing data to visualize.
        title (str) : Title of plot.
        xlabel (str) : X axis title.
        ylabel (str) : y axis title.
        palette : Color palette to create plot using.
```

Returns:

```
"""  
# Create bar plot of feature coefficients as odds.  
fig, ax = plt.subplots(figsize=(10,10))  
  
sns.barplot(x=series.values, y=series.index, palette=palette,  
            ax=ax, orient='h')  
  
ax.set_title(title)  
ax.set_xlabel(xlabel)  
ax.set_ylabel(ylabel);
```

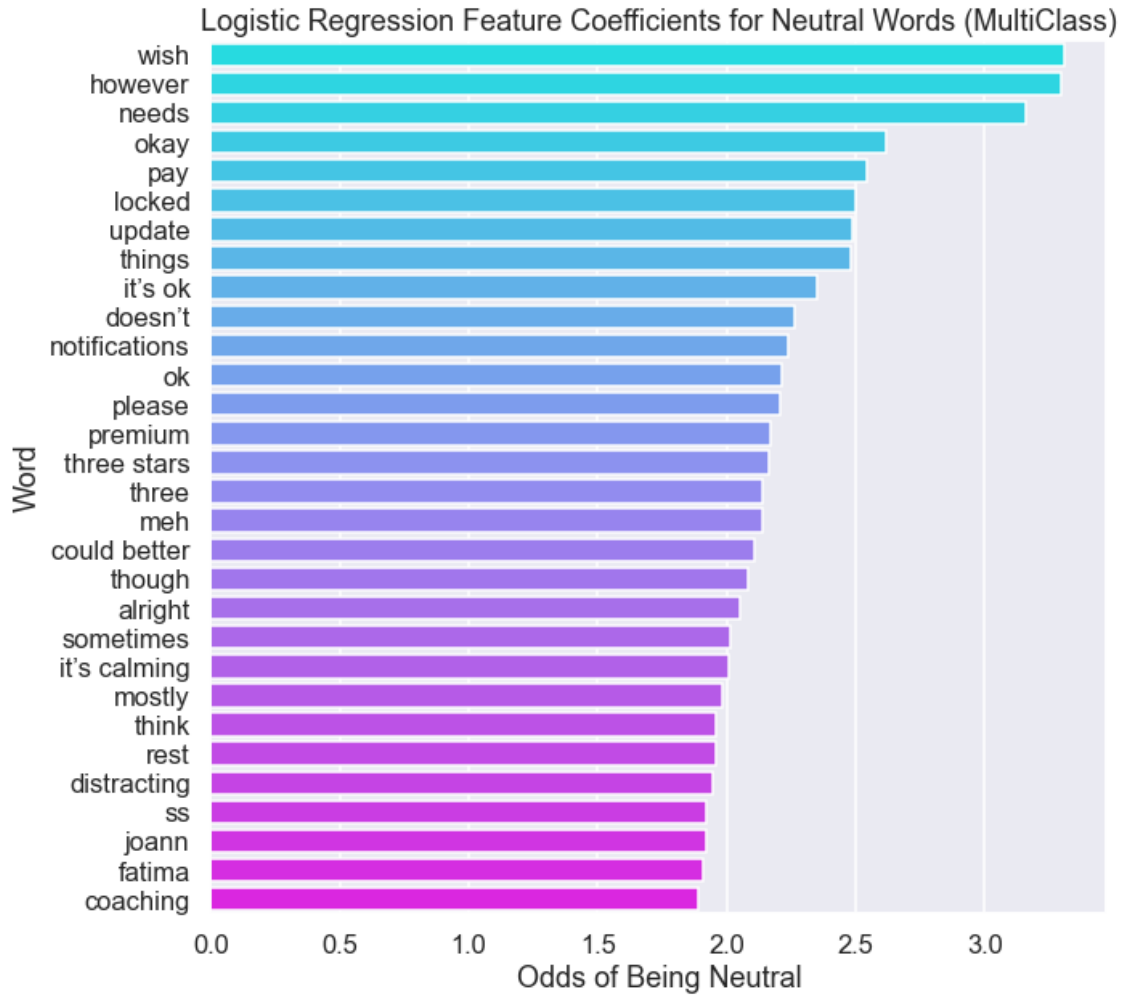
```
[107]: # Set Seaborn style  
sns.set_style('darkgrid')
```

```
[108]: # Create bar plot of feature coefficients for multiclass negative as log odds  
barplot_series(lr_multi_neg,  
               "Logistic Regression Feature Coefficients for Negative Words_  
↳(MultiClass)",  
               "Odds of Being Negative", "Word", palette='cool_r')
```



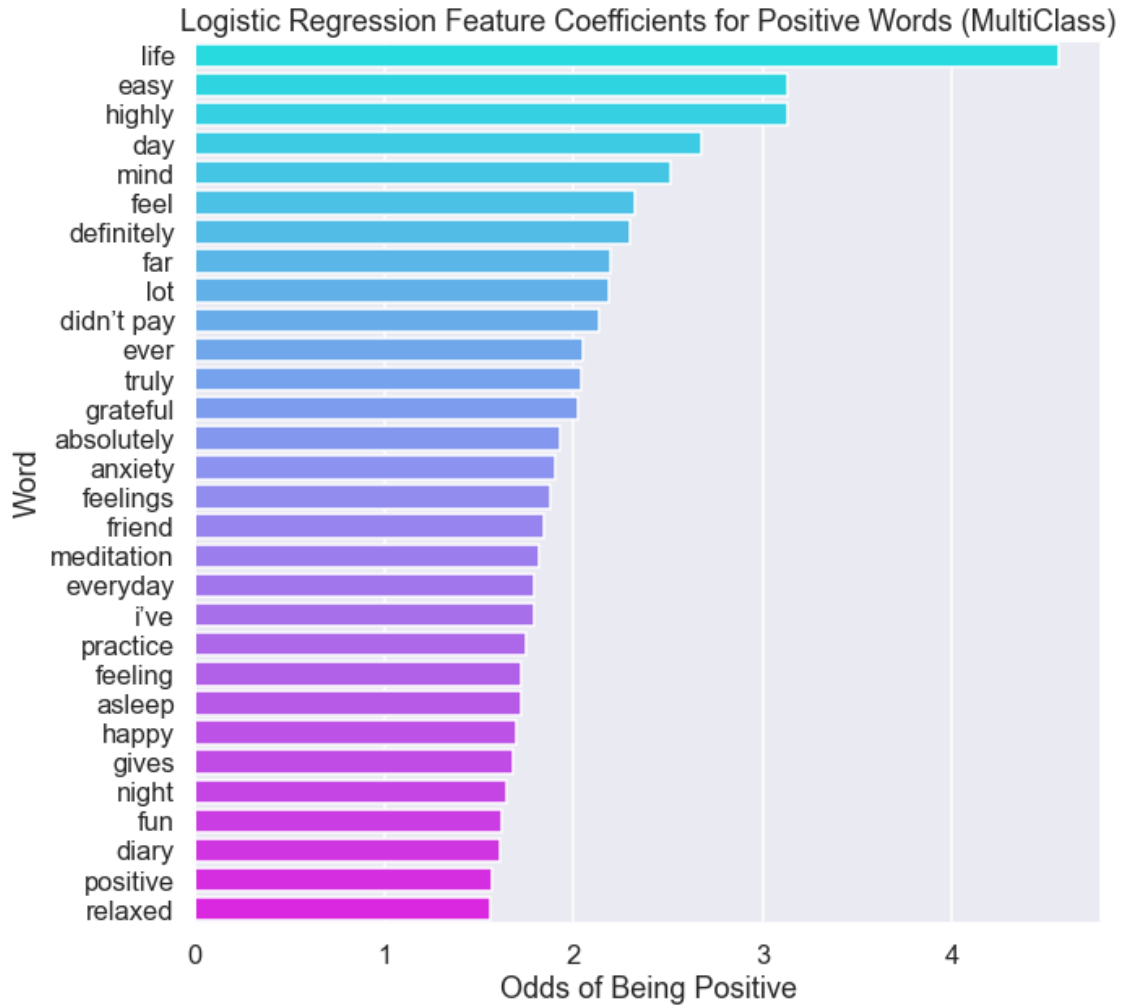
Our Multi-Class negative sentiment coefficients indicate with words such as “pay”, “money”, “charged”, “free”, “charge” and “subscription” that users are mostly unhappy about the costs that are involved in using the app. We can also see from words including “scam”, “refund”, “cancel”, “service”, “account” and “support” that it is likely that a lot of negative reviews are related to customer support.

```
[109]: # Create bar plot of feature coefficients for multiclass neutral as log odds
barplot_series(lr_multi_neut,
               "Logistic Regression Feature Coefficients for Neutral Words_␣
               ↪(MultiClass)",
               "Odds of Being Neutral", "Word")
```



We can see that our model seems to have extracted words that indeed indicate what topics are related to a neutral review. However, this information is not so crucial to our analysis.

```
[110]: # Create bar plot of feature coefficients for multiclass positive as log odds
barplot_series(lr_multi_pos,
               "Logistic Regression Feature Coefficients for Positive Words_
               ↪(MultiClass)",
               "Odds of Being Positive", "Word")
```

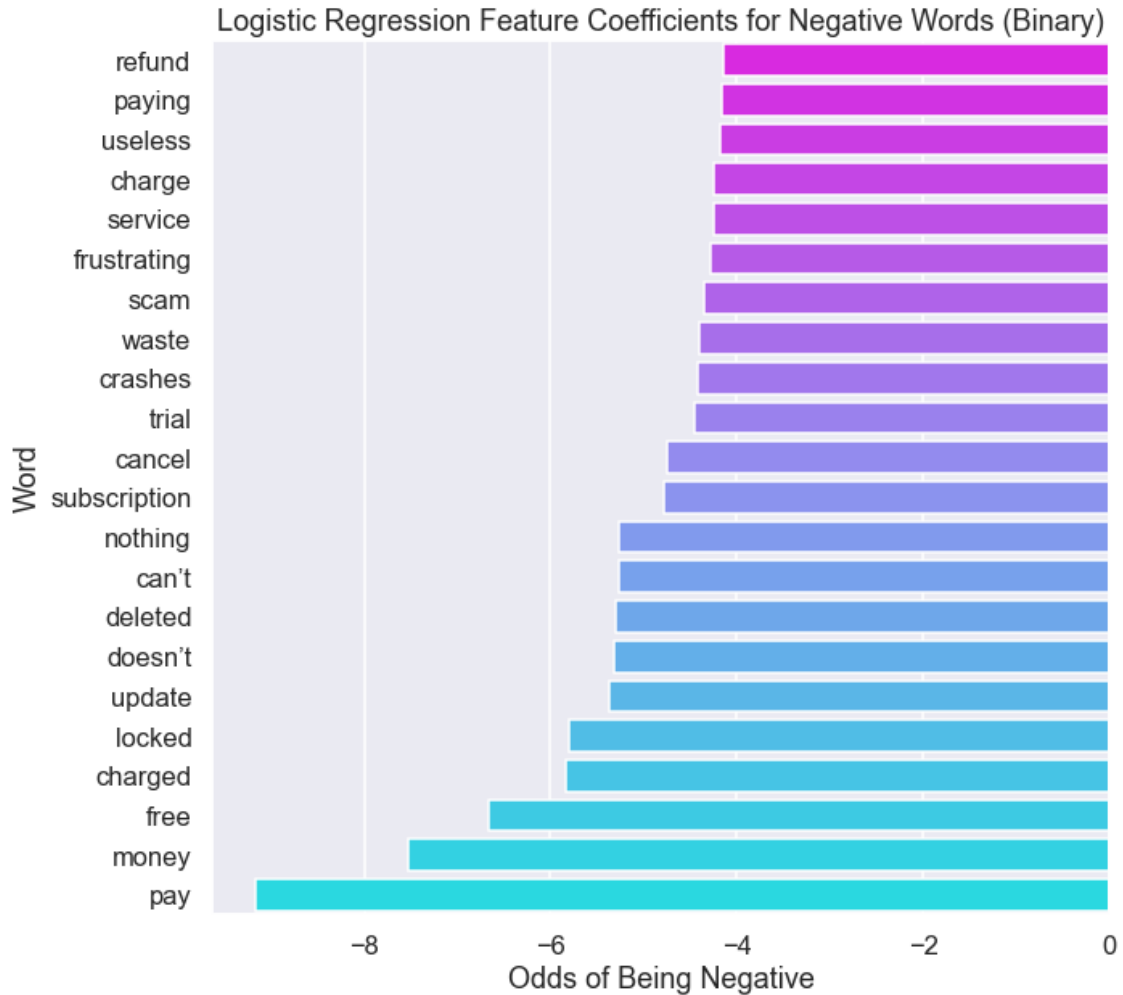


For multi-class positive classification, we can see from the words “asleep” and “night” that sleep functionality is considered a positive element of the app. We also see from the words “easy”, “day”, “everyday” and “diary” that users appreciate ease of use that enables people to use the app regularly.

Now that we have a general idea of what words are associated with negative and positive sentiments, let’s move on to examine whether our binary classification model confirms our findings.

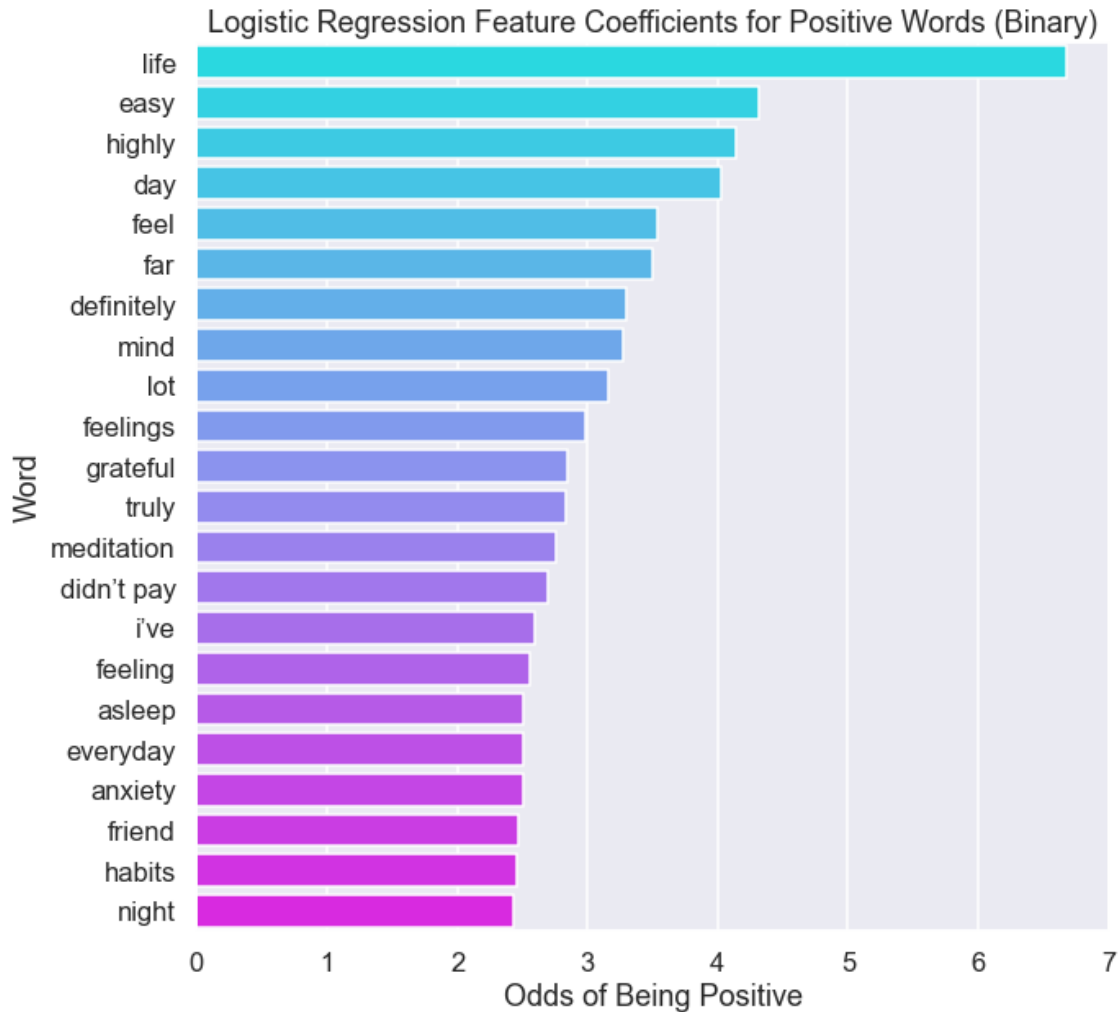
Binary Logistic Regression Coefficients

```
[111]: # Create bar plot of feature coefficients for binary negative as log odds
barplot_series(lr_bin_neg,
               "Logistic Regression Feature Coefficients for Negative Words",
               "(Binary)",
               "Odds of Being Negative", "Word", 'cool_r')
```



Our binary Logistic Regression confirms that payment is a major issue, and through the words “response” and “crashes” further reinforces that customer support is also something to consider while addressing negative reviews.

```
[112]: # Create bar plot of feature coefficients for binary positive as log odds
barplot_series(lr_bin_pos,
               "Logistic Regression Feature Coefficients for Positive Words_
               ↪(Binary)",
               "Odds of Being Positive", "Word")
```



For positive sentiments, we see another key indicator word “habits” which further illustrates that users appreciate being able to make the app a regular part of their routine. We also see the words “night” and “asleep”, which confirm that sleep functionality is considered a highly positive app feature.

Local Interpretable Model-agnostic Explanations (LIME) Finally, let’s use the LIME package to help us better understand exactly how our binary Logistic Regression model is breaking down each review in the process of determining whether it is positive or negative.

```
[113]: # Write function to display LIME text explainer visualizations
def explain_text(idx):
    """
    Creates LIME explainer visualization for a specified review text.

    Args:
        idx (int) : Index number of review text.
```

Returns:

```
"""
c = make_pipeline(vectorizer_bin, best_bin_lr)
class_names=['Negative', 'Positive']
explainer = LimeTextExplainer(class_names=class_names)
exp = explainer.explain_instance(X_bin_test[idx], c.predict_proba,
                                num_features=6)

print('Document id: %d' % idx)
print('Probability(Positive) =', c.predict_proba([X_bin_test[idx]])[0,1])
print('True class: %s' % class_names[y_bin_test[idx]])
exp.show_in_notebook(text=True)
```

```
[114]: # Explain negative review
       explain_text(17187)
```

```
Document id: 17187
Probability(Positive) = 0.020999261127576607
True class: Negative
```

<IPython.core.display.HTML object>

In the above example, we can see that this user is unhappy with how he/she has been charged. This again points to the idea that we need to carefully consider how the user should be charged for the app, as well as what kind of support to provide to address this type of issue.

```
[115]: # Explain positive review
       explain_text(4721)
```

```
Document id: 4721
Probability(Positive) = 0.9423697140038891
True class: Positive
```

<IPython.core.display.HTML object>

In this positive example, we can see that there are some words such as “pain”, “restlessness” and “struggle” that seem negative. However, our model has determined that there are other words that have higher importance, including “mind”, “sleeping”, “forward” and “asleep” that carry much more weight in classifying this review as a positive one.

Now, let’s take a look at a review that our model was unable to predict correctly.

```
[116]: # Explain incorrect classification
       explain_text(3212)
```

```
Document id: 3212
Probability(Positive) = 0.5265458013611568
True class: Negative
```

<IPython.core.display.HTML object>

When we look at the text, it is obvious to us that the word “stupid” is an indicator of negative sentiment, but our trained model was unable to pick up on this word, because we intentionally removed it during the stopword removal phase. The reason behind this was to eliminate any obvious indicators of sentiment in order to allow the model to learn what specific topics were considered negative or positive.

Although our model correctly determines that the words “free” and “premium” indicate a negative sentiment, our model also considers any mention of “fall” and “asleep” to be positive. The prediction probabilities indicate that the model was unable to easily determine that this was a negative review.

This example shows us that although our model might make occasional errors in correctly predicting the overall sentiment, it does give us a good idea of what words and topics are considered positive or negative.

8 CONCLUSIONS & RECOMMENDATIONS

Now that we have thoroughly analyzed the EDA and models of mental health app reviews and ratings throughout the course of this notebook, we can return to the questions outlined in the Business Problem phase, which are:

1. What do users like or dislike about apps currently available on the iTunes App Store?
2. How can we use this information to develop a strategy for building a new mental health app that can compete with apps that have already seen success?

First, we now know that the following topics are highly correlated with negative sentiment: > - Payment and costs > - Technical and customer support

These topics are considered highly associated with positive sentiment: > - Sleep-aid functionality > - Ease-of-use and regular use

Using these insights, we now have an idea of what elements to focus on when building a new mental health app to compete with other apps already on the market. My recommendations are as follows: > 1. Since payment and costs seem to be a prevalent issue in negative reviews, app pricing and subscription fees needs to be well thought-out. Many reviews mentioned dissatisfaction in the “free” aspect of the app, indicating that they may have felt misled. We should address this issue by specifying at the time of download about what features of the app are free and what features are reserved for premium members. > 2. Allocate sufficient resources toward thoroughly training the customer support team. Users expressed dissatisfaction in when their app crashed, as well as not having their payment issues properly resolved. We should make sure that customer support is trained to proficiently diagnose any technical issues as well as any payment-related issues users might experience. > 3. Invest in R&D of the sleep-aid functionality of the app. Users associated sleep-related words with positive sentiment, and this illustrates that an app’s ability to become an effective sleeping aid can be a determinant factor of its overall effect on the user’s mental health. > 4. Thoroughly develop the interface of the app to make it pleasant and easy to use. Users are likely to associate ease-of-use and the ability to make using the app a regular habit with positive sentiment, so the UI/UX designers will want to address this accordingly.

Some limitations and ideas for further analysis include: > - We currently do not have any pricing data of apps. Further research and analysis could be done regarding pricing of subscription and app costs to determine what the best pricing scheme would be. > - Our data is currently limited

to the reviews and ratings on the Apple App Store. Collecting and analyzing app reviews from the Google Play Store to examine how they compare would help us get a well-rounded idea of the sentiments of all mobile app users, as opposed to just Apple users. > - Possible implementation of Latent Dirichlet Allocation topic modeling to extract deeper insights on what topics could be found in negative vs positive reviews.

[]: