Topic Modeling- Meditation App Reviews

Sai Varadharajan
LIS 501

Motivation/Intro

- Meditation apps have become more and more popular these days, with people becoming interested in the benefits of meditation
- I am curious about what people have to say regarding these apps

Objectives

- Perform Topic Modelling using BERtopic and create visualizations
- Explore the different visualizations created based on different hyperparameter tunings

Background

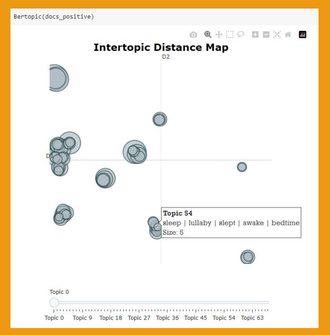
- I will be performing text mining analysis on positive and negative reviews of these apps to figure out what topics frequently come up
- I will also illustrate the effects of hyperparameter tuning

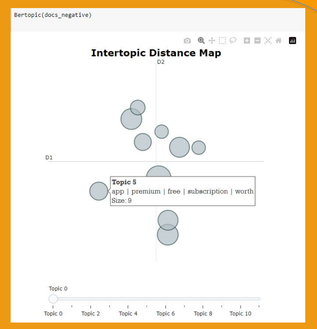
Methods

- I have separated the dataset into two, one for positive reviews (with a rating higher than 3), and one with negative reviews (with a rating less than 3).
- I then performed topic modeling on the comments of each dataset using BERtopic and then created visualizations representing these clusters of topics.
- I created different visualizations based on different values for the hyperparameters, namely the n_components and n_neighbors for the positive and negative reviews respectively.
- I compared and contrasted these visualizations to see which parameter values enable the best topic models.

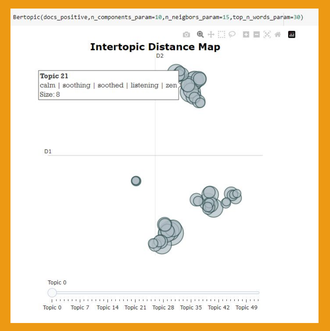
Results

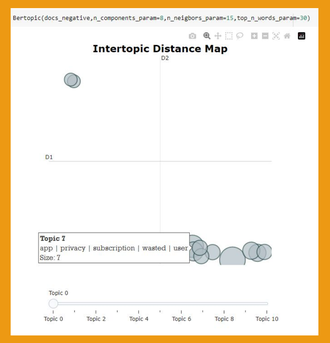
Default Values



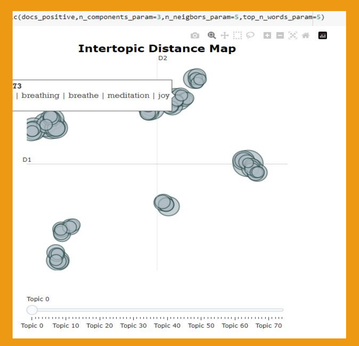


Increasing params





Decreasing Params



Conclusion/Findings

- We can see that increasing n_components too much leads to BERtopic having a hard time clustering the high-dimensional embeddings, while lowering the value too much causes not sufficient info for proper clusters
- op_n_words: We can see that increasing n_words too much creates less coherent topic formation, while having a value that's too low causes less informative/representative topics
- ncreasing n_neigbors creates a more global view of the embedded structure and larger clusters being created, while decreasing it leads to a more local view and fewer clusters.