Abstract
The start of the new year is a time when individuals make resolutions to better their lives. All resolutions have merit yet require different types of support to ensure lasting changes can be made. Using tweets scraped from Twitter, I built unsupervised learning models that identify the primary topics of interest related to New Year's resolutions and the general sentiment related to these tweets. The topics identified and the sentiment of these topics suggest there is *not* a need for more content surrounding goal planning. Rather, positive content regarding alcohol consumption and diet are areas where additional support could be provided.

Design
There are many types of New Year's resolutions and each requires a different style of support. With the majority of individuals failing to keep their resolutions even after a single month, the goal of this project was to identify topics of interest related to New Year's resolution Twitter posts and the sentiment of these posts. This analysis would then help wellness companies create relevant and supportive content for individuals to help them make lasting changes to improve their lives.

Data
Nearly 17,000 tweets from 1/5/22 – 1/19/22 were scraped from Twitter by searching for the terms "New Years Resolution", "2022 Resolution", "#NewYearsResolution", and "#NewYearResolution". The searches included only tweets in English and excluded retweets that did not have a user comment. In addition to the text of the tweets, the number of favorites and retweets for each tweet were also collected.

Algorithms
Exploratory data analysis was performed and initial topic modeling was performed/explored using CountVectorizer and TfidfVectorizer as well as LSA, NMF, and LDA. Using some of the terms identified in the exploratory analysis, CorEx was used to finalize the topics.

Sentiment analysis was performed across the entire dataset and on each of the individual topics. The sentiment of the most popular tweets was also explored (most favorited, most retweeted).

Figures were created to highlight the top terms associated with each topic (word clouds) and to highlight the different terms used between positive and negative tweets.

Tools
The tweets were obtained by scraping Twitter using Tweepy. Exploratory data analysis and unsupervised model building were performed with Pandas, unicodedata, spaCy, gensim, scikit-learn, and corextopic. The sentiment analysis was performed with VADER. Figures were created with Matplotlib, Scattertext, and WordCloud.

Communication
Slides and visuals were presented and posted on my personal GitHub page.