An Analysis of Twitter Posts During the Covid-19 Pandemic

Can we predict public sentiments as new variants emerge?

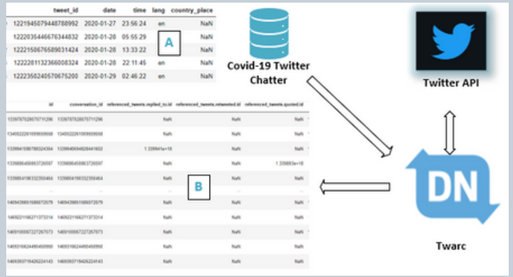
Team 68 : Arianna Cooper, Yoseph Dawit, Funke Olaleye, Suzan Pham

The Web Application

- Data Frame - Explore data frame, observe correlations, & create scatterplots
- Heat Map - Vizualise the sentiments across each hour
- Sentiment - Examine the various sentiments trends along the variants
- Prediction - Type in a sentence to see the sentiment analysis
- Word Cloud - Vizualize the most common words used in posts

Background

Data analysis was performed on the numerical features related to Twitter account users' and their posts. The module VaderSentiment, was utilized in performing Natural Language Processing (NLP) Sentiment analysis. Once patterns and trends were uncovered, a web application was created highlighting the findings through the following: The Dataframe, Sentiment, Heat Map, Word Cloud, and Prediction. A user-friendly application was created by using an open source package called Streamlit.

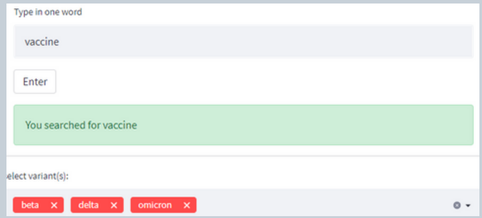


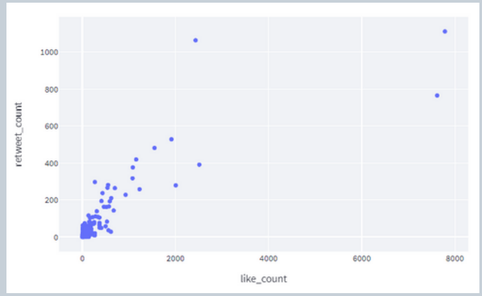
Data Collection and Pre-Processing

- Source - Covid-19 Twitter Chatter dataset from Zenodo and Tweeter API
- Collection -Utilized Twarc2 package to transform list of tweet_ids to full tweets by hydration and selected posts within a 10-day window of each variant (Beta, Delta, & Omicron)
- Tabular Pre-processing : Extracted each hour, renamed columns, created variant name by date, & parse data types
- NLP Text Pre-processing : Applied lowercase, removed punctuations, URLs, & stop words, tokenized and lemmatized tweet text

Data Frame

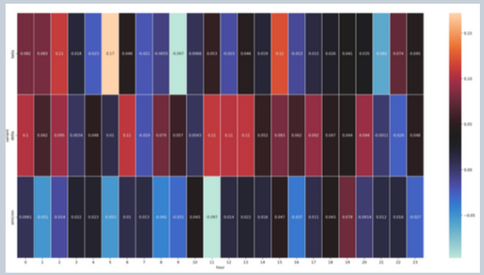
- The entire data frame is displayed by default
- User can adjust how many rows to show
- Filter options: Variant (Beta, Delta, & Omicron) user-selected columns, specific date range,
- Posts can be filted by user-input keyword search
- Correlation map between numerical columns
- Scatterplots based on user-selected columns





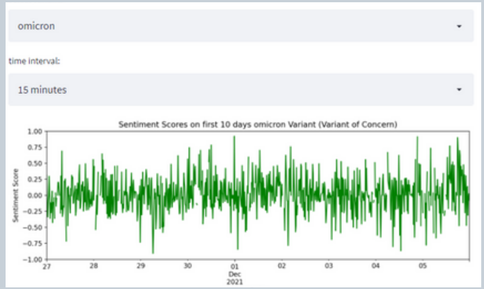
Heat Map

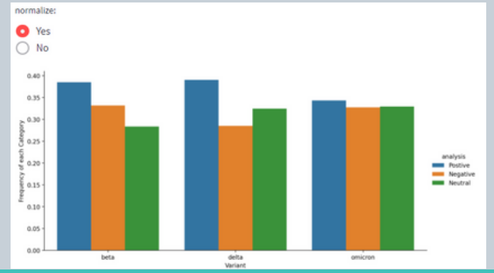
- Sentiment scores of all text, aggregated into one value (mean), across hour (x-axis) & variant (y-axis)
- Radio button 'no' allows the user to input a text for filtering and display the heat map on the subset of sentiment scores



Sentiment

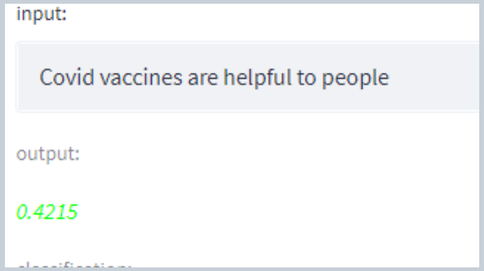
- A bar-plot highlighting sentiment across each variant (Beta, Delta, & Omicron
- Trend - More shares of positive sentiment for the Beta & Delta variants
- Time series aggregated by selected time intervals & selection of variant





Prediction

- Used VaderSentiment module to extract sentiment scores
- Classify scores into 3 categories: Positive (green), Negative (red), and Neutral (gray)
- Sentiment analysis based on user-input text, where score & classification are returned
- Score ranges from -1 to 1 with the following ranges: Neutral (-0.5, 0.5), Positive (0.5, 1), & Negative (-1, -0.5)



Word Cloud

- The more frequent the word was used, the larger the size of the word display
- Each text has a sentiment score highlighted by color:
- Red for negative, green for positive, & gray for neutral
- Slider allows user to define the number of words to display