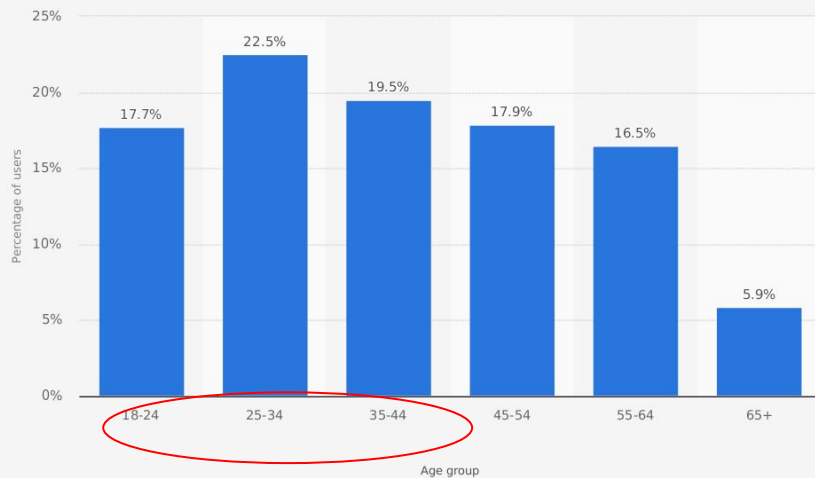# Does VADER sentiment analysis produce useful data for suicide prevention outreach on Twitter?

A Case Study

The Data We Know

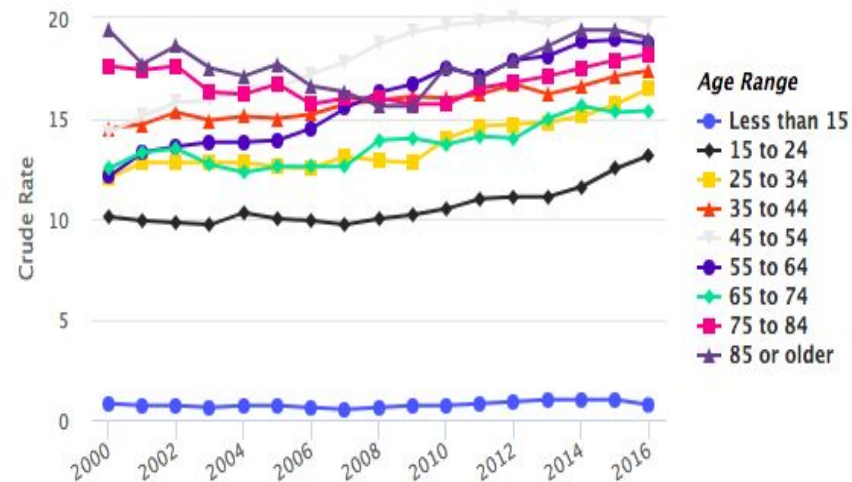
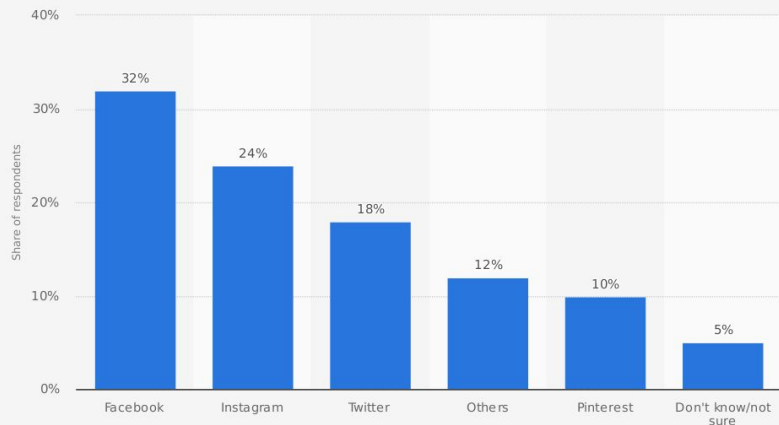Distribution of Twitter users in the United States as of December 2016, by age group

Suicide Rates by Age from 2000 to 2016

Which Data Could Help National and Local Orgs?

Best social media platforms for influence marketing according to influencers in the United States as of July 2016

- Facebook: 32%
- Instagram: 24%
- Twitter: 18%
- Others: 12%
- Pinterest: 10%
- Don't know/not sure: 5%

Sources
eMarketer; SheSpeaks
© Statista 2018

Additional Information:
United States; SheSpeaks; July 2016; 347 respondents; among influencers with an active blog and who were act

American Foundation for Suicide Prevention
@afspnational

Preventing suicide through research, advocacy, and education. Account not monitored 24/7. Thoughts of suicide? Call: 800-273-8255 / Text: 741741

⊙ USA
🔗 afsp.org
📅 Joined January 2009

SFSuicidePrevention
@SFSuicide

We're a 24-hour suicide prevention hotline available to anyone in emotional crisis. If you need help or someone you know is in crisis, please call 415.781.0500

⊙ San Francisco
🔗 sfsuicide.org
📅 Joined October 2011

Project Overview & Goals

Goal: To create actionable data (maps, keywords, trends) for suicide and crisis prevention organizations to target ads to Twitter users experiencing emotional distress.

Data source: English language tweets (as selected by Twitter) from several US States

Cleaning & parsing data: Created a list of words and phrases that indicate emotional stress and distress, cleaned raw csv pulled from Twitter filtering by distress list, human filter; created dataframes, re-encoded csvs to UTF-8.

Sentiment Analysis Analyzer: VADER

Visualization: WordCloud and statistical graphics

Assumptions & Caveats

Sample size and time series are not large enough to make definitive judgments from data.

VADER does not pick up on sarcasm.

(Ex: 'Warren Beatty has never been so entertaining.' compound: 0.5777, neg: 0.0, neu: 0.616, pos: 0.384)

VADER is not great with simultaneously reading polarity and strength and slang.

(Ex: 'VADER sentiment analysis is the s***.' compound: 0.6124, neg: 0.0, neu: 0.556, pos: 0.444)

We don't know how time and day are related to Twitter usage or sentiment. There may be a day/time that is more reflective of suicidal sentiments and better suited for targeted ads.

Data Gathering and Clean-up

- Twitter is a poor source of large data sets
  - Limited volume per call
  - No historical data sets for free
- Twitter data requires a lot of cleaning
  - Not uniform data
  - Significant filtering required
- Public data is already presented in summary format
  - Scraping not feasible
  - Purpose-built summaries
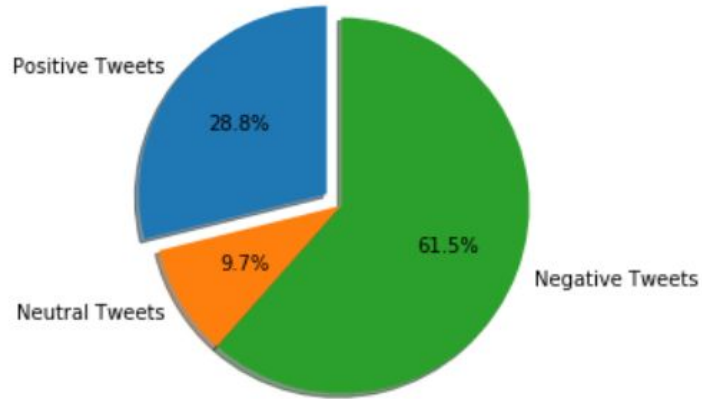
Exploration of Data

- Researched Trigger Words
- Filtered Tweets
- Measured Sentiment
- Discovered our hypothesis was incorrect
  - Led us to explore why…

"The word ibuprofen was 16 times more likely to predict the person texting would need emergency services than the word suicide."
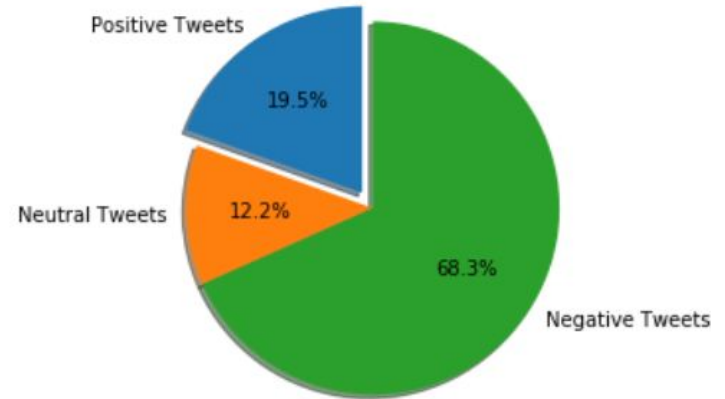
- https://mashable.com/2017/03/09/artificial-intelligence-suicide-risk/#HikouaxJEmqi

Exploration and Analysis

Sentiment Analysis on Suicidal Tweets - 2009 1.6m Tweets USA

Positive Tweets 28.8%
Negative Tweets 61.5%
Neutral Tweets 9.7%

Sentiment Analysis on Suicidal Tweets - California

Positive Tweets 19.5%
Neutral Tweets 12.2%
Negative Tweets 68.3%

Word Analysis

Let's Go to JUPYTER NOTEBOOK!

Conclusions and Implications

Sentiment analysis of tweets with "suicide trigger words" may not produce useful information for local and national suicide prevention twitter outreach . *Derived from 4 pie and 1 bar chart.*

- Twitter may not be the preferred social medium for expressing suicidal inclinations.
- The data is likely inconclusive due to the small time frame.
- There may be a better sentiment analyzer (textblob, SlangSD, etc.)
- There is evidence that specific words, and not sentiments, is a better indicator (ex. ibuprofen) of potential self-harm

References

Gilbert, CJ Hutto Eric. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (2018/04/16) http://comp. social. gatech. edu/papers/icwsm14. vader. hutto. pdf*. 2014.

American Foundation for Suicide Prevention. *Available at (2018/04/16)* https://afsp.org/about-suicide/suicide-statistics/

Statista. "Distribution of Twitter users in the United States as of December 2016, by age group" *Available at (2018/04/16)* https://www.statista.com/statistics/192703/age-distribution-of-users-on-twitter-in-the-united-states/

Statista. "Best social media platforms for influence marketing according to influencers in the United States as of July 2016." *Available at (2018/04/16)* https://www.statista.com/statistics/617525/best-social-media-for-influence-marketing-per-influencers/

Crisis Trends. " Co-occuring issues. Texters in all states experiencing anxiety/stress." www.crisistrends.org

Thank you