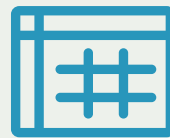# Sentiment Analysis of Depression in Tweets

Jordana Tepper

Agenda

BUSINESS PROBLEM

DATA UNDERSTANDING

DATA ANALYSIS

MODELING

LIMITATIONS

NEXT STEPS

Business Problem

GLOBAL DEPRESSION

280 million

UNITED STATES DEPRESSION RATE

1 in 10 people

UNITED STATES SOCIAL MEDIA USAGE

7 in 10 people

GLOBAL SOCIAL MEDIA USAGE

59 percentage

Data Understanding
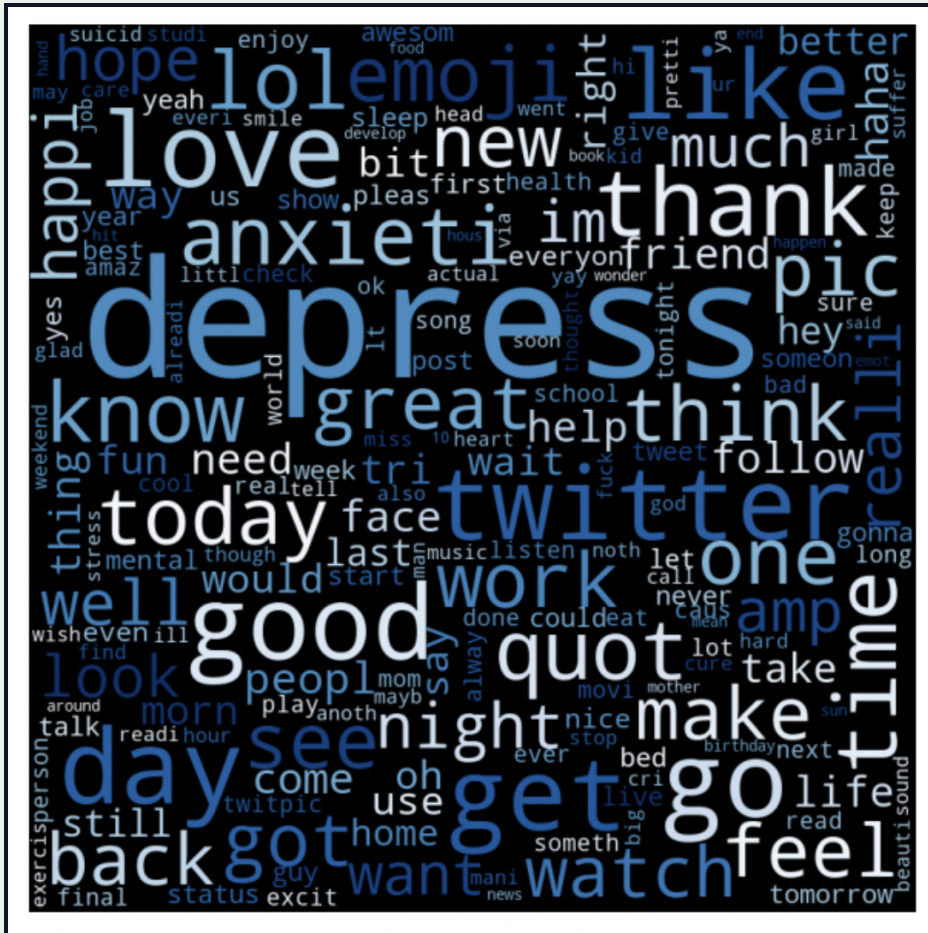
1 **DATA SOURCE: KAGGLE**

2 **DATA SET SIZE: 10,313 ROWS, 3 COLUMNS**

3 **PREDICTOR VARIABLE: TWEETS**

4 **TARGET VARIABLE: INDICATION OF DEPRESSION (0 OR 1)**

Data Analysis

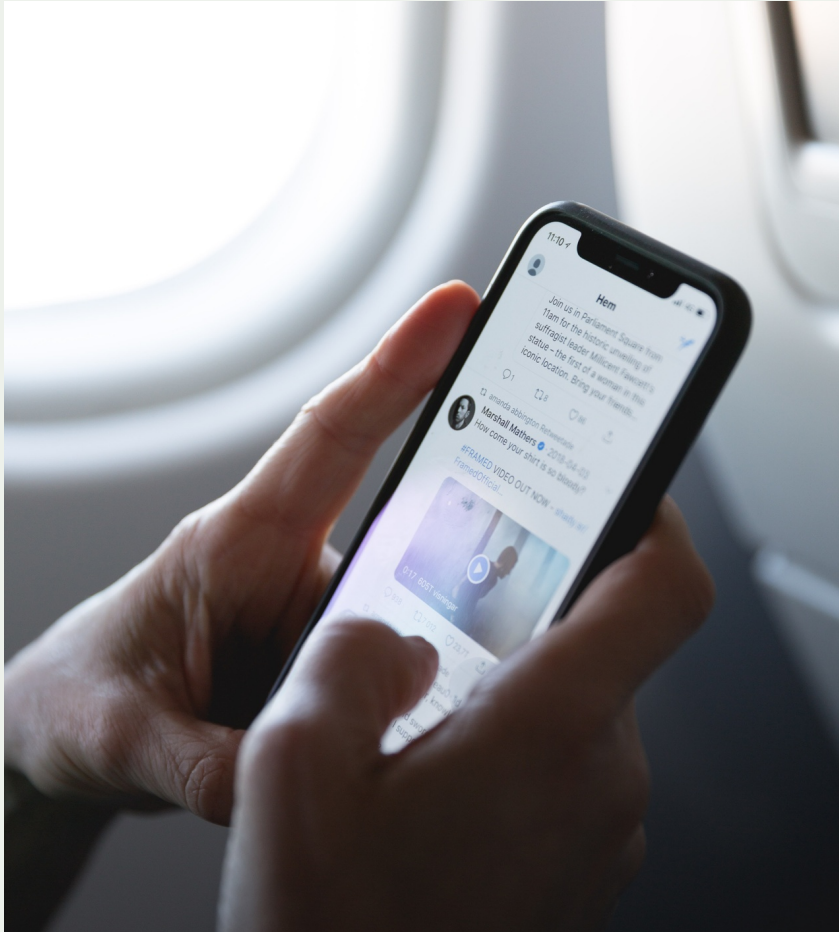## WORDS WITH THE HIGHEST FREQUENCY IN THE DATA SET (STEMMED)

1 | DEPRESS

2 | GO

3 | GOOD

4 | LOVE

5 | DAY

Limitations

1 | THE MODEL FAILS TO UNDERSTAND SLANG WORDS SUCH AS 'EMO' AND 'DEPRO'

2 | THE MODEL ALSO FAILS TO PROPERLY INTERPRET NEGATION IN STATEMENTS SUCH AS, "I AM NOT DEPRESSED" OR "I AM NOT OKAY"

3 | THE DATA SET CONTAINS ONLY 10,313 TWEETS MEANING THAT THE MODEL IS LIMITED AND CANNOT CAPTURE EVERY PATTERN OF TEXT/LANGUAGE THAT EXISTS ONLINE.

Next Steps

1 | **ACQUIRE A LARGER DATA SET**

2 | **ACQUIRE MORE DATA THAT ACCOUNTS FOR SLANG WORDS**

3 | **DEVELOP MODELS THAT INCORPORATES OTHER LANGUAGES**

Contact Info

@ jtepper724@gmail.com

🌐 github.com/jordanate

in linkedin.com/in/jordana-tepper/