



TRUMP TWITTER ARCHIVE

# TRUMP TWITTER ARCHIVE

“...the Trump tweets are being preserved for posterity, so future generations can revel in the president's peculiar and unprecedented use of Twitter as an art form and governing tool.”

-source theVerge



# TRUMP TWITTER ARCHIVE

The trumptwitterarchive is a collection of his every tweet, retweet, and favorite, and this project plans to load and predict engagement levels based on text, hastags, sentiment, and other feature columns. Analysis on data will start after Trump's acceptance as the republican nominee at National Republican Convention on July 18, 2017.

Goals:

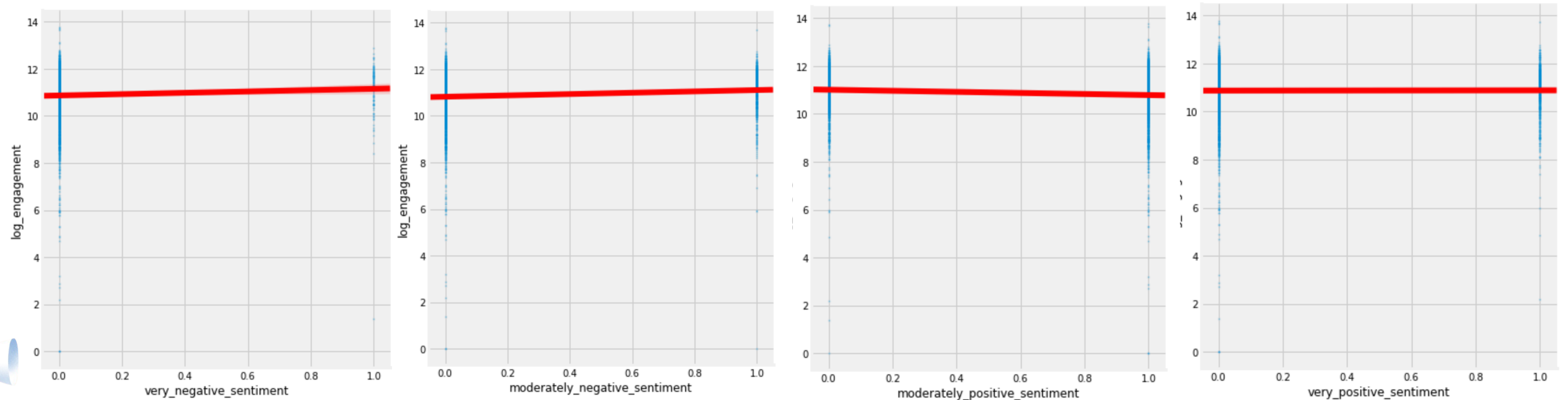
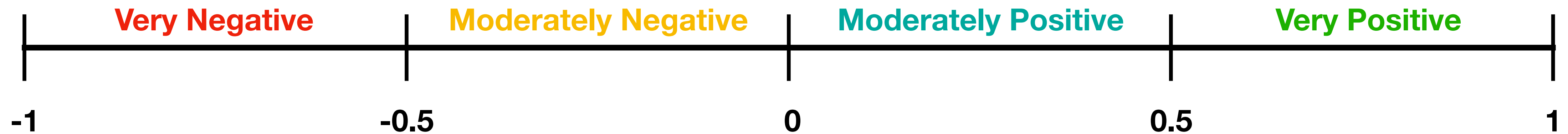
Predict twitter engagement based on the sentiment of Trump's tweets

1. Sentiment Analysis
2. Linear Regression Results
3. NLP to review Trump's Vocabulary

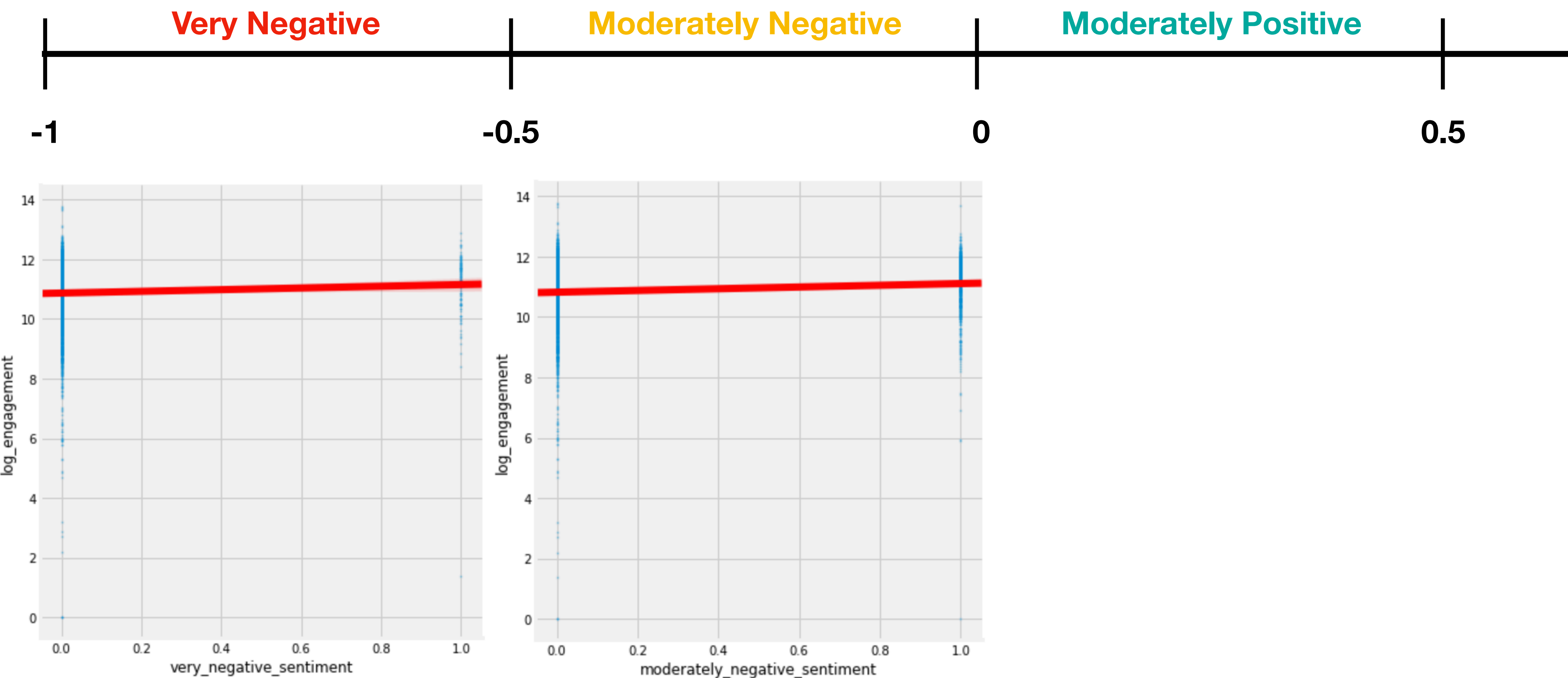


# SENTIMENT ANALYSIS: Methodology

1. Used the built `Textblob().sentiment.polarity` function
2. Created Sentiment Bins
3. Used Sentiment Dummy Variables in the Linear Regression Model



# SENTIMENT ANALYSIS: Negative Sentiment Scatter Plot



# SENTIMENT ANALYSIS: Tweets with Highest Engagement



“My warmest condolences and sympathies to the victims and families of the terrible Las Vegas shooting. God bless you!” 10/2/2017

“We must keep "evil" out of our country!” 2/3/2017

“Terrible! Just found out that Obama had my "wires tapped" in Trump Tower just before the victory. Nothing found. This is McCarthyism!” 3/4/2017



# SENTIMENT ANALYSIS: Tweets with Highest Engagement



“My prayers and condolences to the families of the victims of the terrible Florida shooting. No child teacher or anyone else should ever feel unsafe in an American school.” 2/14/2018

“...Senators should focus their energies on ISIS illegal immigration and border security instead of always looking to start World War III.” 1/29/2017

“A disgraceful verdict in the Kate Steinle case! No wonder the people of our Country are so angry with Illegal Immigration.” 12/1/2017





# LINEAR REGRESSION: Factors

The Linear Regression model predicts:

$\log(\text{engagement})$

On the following factors:

is\_retweet

is\_afternoon

is\_evening

is\_latenight

moderately\_positive\_sentiment

moderately\_negative\_sentiment

very\_negative\_sentiment





# LINEAR REGRESSION: Results

## The Linear Regression Results:

MAE 0.58

MSE: 0.81

RMSE 0.90

intercept: 11.22

is\_retweet: -1.97

is\_afternoon: -0.26

is\_evening: -0.24

is\_latenight: -0.28

moderately\_positive\_sentiment: -0.08

moderately\_negative\_sentiment: 0.10

very\_negative\_sentiment: 0.23

## Model of Means

MAE: 0.78

MSE: 1.18

RMSE: 1.09



# LINEAR REGRESSION: Results Regularization w/Lasso

## The Linear Regression Results:

MAE 0.58

MSE: 0.81

RMSE 0.90

intercept: 11.22

is\_retweet: -1.97

is\_afternoon: -0.26

is\_evening: -0.24

~~is\_latenight: 0.28~~

moderately\_positive\_sentiment: -0.08

moderately\_negative\_sentiment: 0.10

~~very\_negative\_sentiment: 0.23~~

## Model of Means

MAE: 0.78

MSE: 1.18

RMSE: 1.09



# Natural Language Processing: CountVectorizer

CountVectorizer(min\_df=5, max\_features=5000 ,stop\_words='english', ngram\_range=(1,3))

Unique Phrases/Words = 2641

Great 822	Thank 447	People 368	Trump 348	America 342
Hillary 301	President 268	Today 266	News 257	Clinton 250
Big 244	Country 239	Make 225	Time 212	Fake 206
realdonaldtrump 201	Jobs 200	New 198	American 197	Tax 192
Media 190	Fake News 175	Vote 152	Crooked 151	Make America Great 120



# TRUMP TWITTER ARCHIVE

The End

