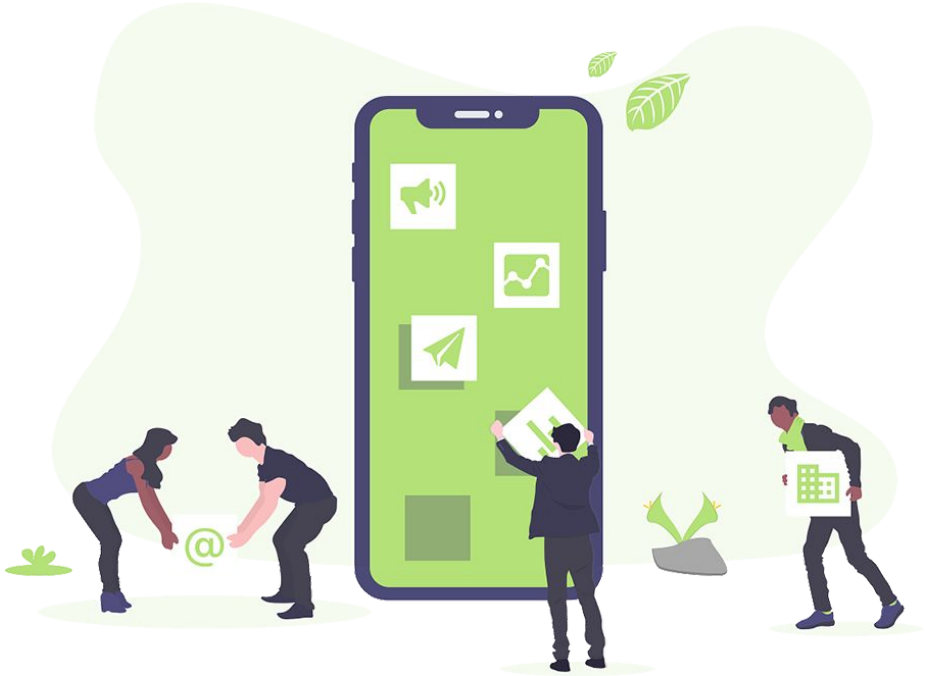
Phase 4 Project
Leah Pope

NLP for Classifying Consumer Tweets

81720-ONL-DS-FT
12/11/2020



Hello!

I’m Leah Pope

I’m a Flatiron Data Science Student and Software Engineer

https://github.com/lspope/

“How are consumers feeling (Tweeting) about our brands and events?

- Tech Companies & Conferences

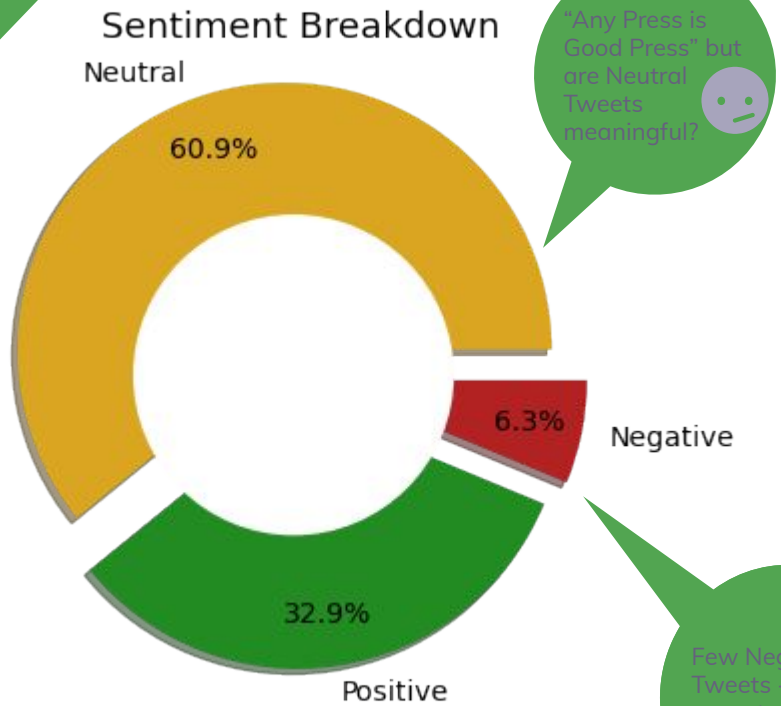
Let’s Find Out

EDA on Raw Data and Corpus

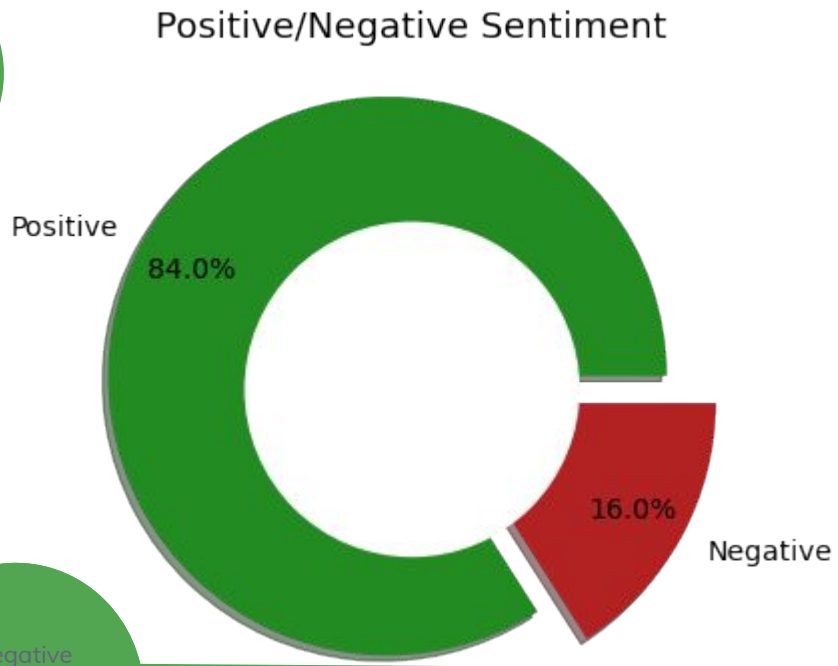
Business Questions

1. Sentiment breakdown?
2. Equally represented?
3. Brand insights?

1 Sentiment breakdown?



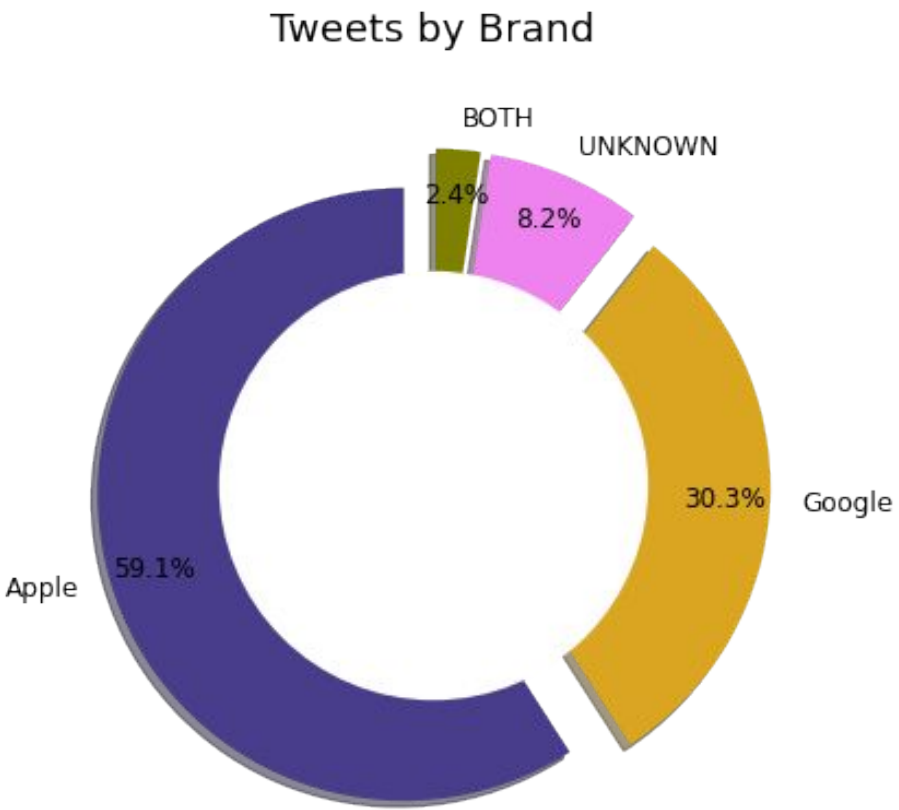
"Any Press is Good Press" but are Neutral Tweets meaningful?



Few Negative Tweets - brands seem to be satisfying.

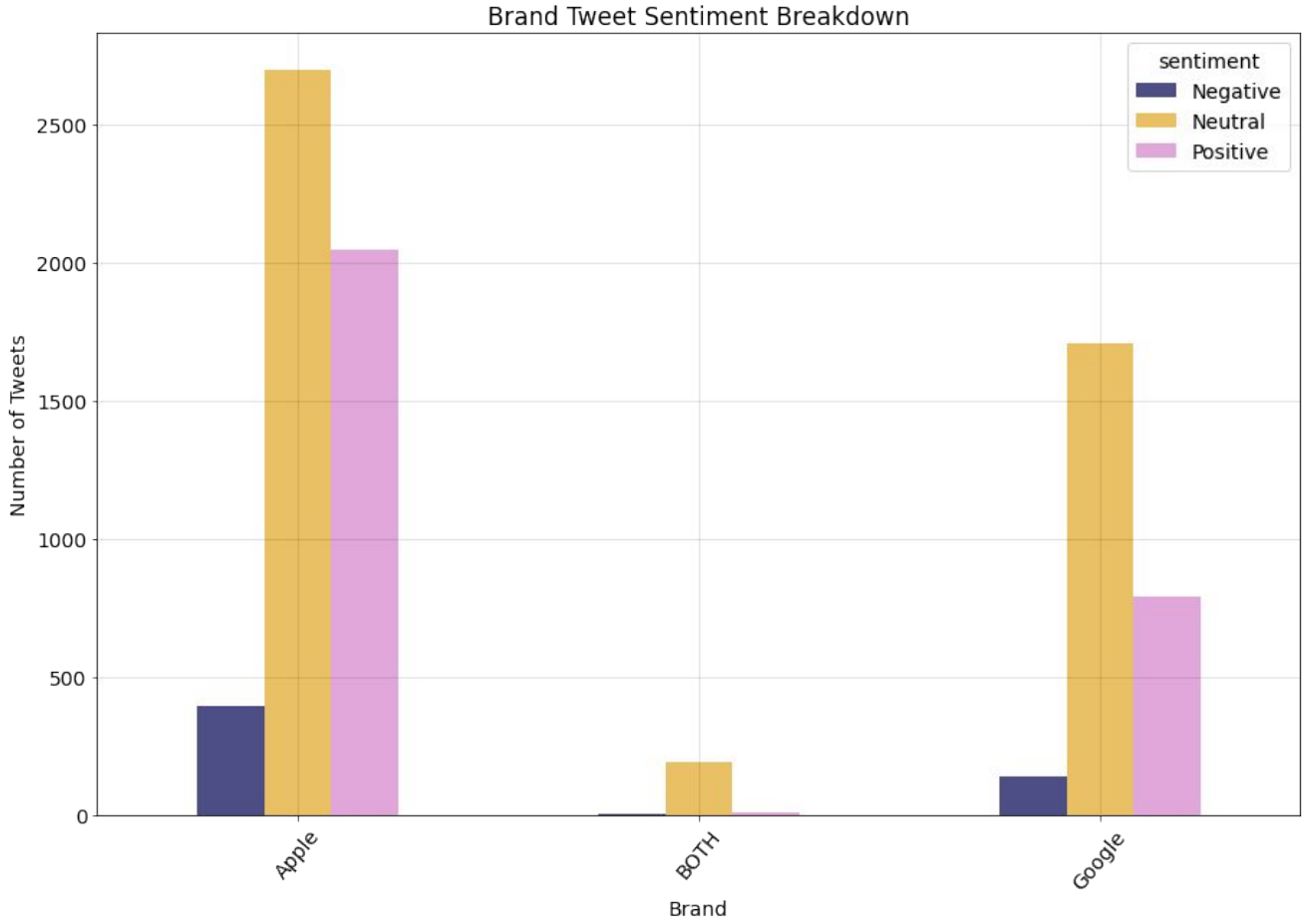
2 Equally Represented?

- Apple leads
  - ~twice the tweets as Google
- Google trails
- Few mentions of both in same tweet
- What about Unknowns?
- Tweets from SXSW
  - Event-goers an Apple crowd?

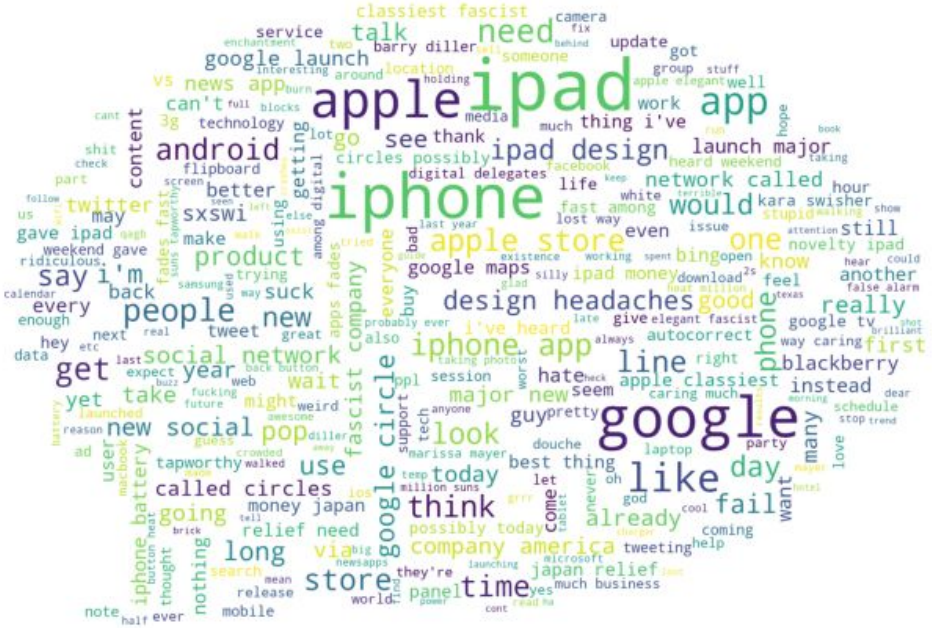


3 Brand insights?

- Few co-mentions may show “Two Camps”
- Similar patterns
  - Pos **>** Neg
  - **++** Neutral
  - Neg:Pos Ratios
    Apple-19%
    Google-18%



What are people tweeting about?


apple
ipad
iphone
google
android
need
app
ipad design
apple store
design headaches
social network
iphone app
line
blackberry
like
think
fail
people
new
social
circle
fascist
company
google circle
launch major
network called
would
still
another
really
first
instead
today
day
long
store
time
panel
america
relief
japan
tweeting
service
talk
update
google launch
news app
can't
work
thing i've
go
see
thank
circles possibly
digital delegates
life
hour
kara swisher
novelty ipad
bing
google maps
ipad money
feel
good
google tv
give
autocorrect
way caring
apple classiest
hate
wait
take
yet
major new
guy
look
pop
use
called circles
going
relief need
via
company
tweeting
japan relief
much business
sxsw
twitter
product
suck
i'm
say
every
tweet
get
year
session
seem
pretty
douche
marissa mayer
best thing
money japan
possibly today
coming
help
microsoft
they're
mobile
release
content
technology
flipboard
better
make
weekend gave
gave ipad
ridiculous.
back
next
hey
data
enough
calendar
trying
samsung
apps fades
everyone
buy
bad
probably ever
i've heard
late
elegant fascist
also
making photos
ppl
works
support
tech
anyone
laptop
party
many
love
want
come
let
oh
god
cool
already
coming
million suns
tapworthy
walked
crowded
battery
iphone
nothing
thought
search
mean
newsapps
launching
much
yes
cont
note
half
ever
brick
big
heard weekend
facebook
white
lost way
fast
among
even
issue
stupid
bing
open
know
one
novelty ipad
another
false alarm
google tv
brilliant
first
blackberry
instead
schedule
stop
trend
dear
morning
caring much
apple classiest
seem
hate
session
ppl
works
right
line
phone
good
feel
texas
download
2s
silly
ipad money
existence
working
spent
last year
bad
everyone
buy
apps fades
company
fascist
circle
google circle
support
tech
anyone
laptop
party
many
love
want
come
let
oh
god
cool
already
coming
million suns
tapworthy
walked
crowded
battery
iphone
nothing
thought
search
mean
newsapps
launching
much
yes
cont
note
half
ever
brick
big
microsoft
they're
mobile
release
world
panel
time
store
via
long
relief need
going
money japan
called circles
use
tapworthy
pop
new social
get
yet
take
wait
year
social network
people
new
tweet
great
every
say
i'm
suck
product
make
weekend gave
gave ipad
ridiculous.
back
next
hey
data
enough
calendar
trying
samsung
apps fades
everyone
buy
bad
probably ever
i've heard
late
elegant fascist
also
making photos
ppl
works

Tweet Classifier

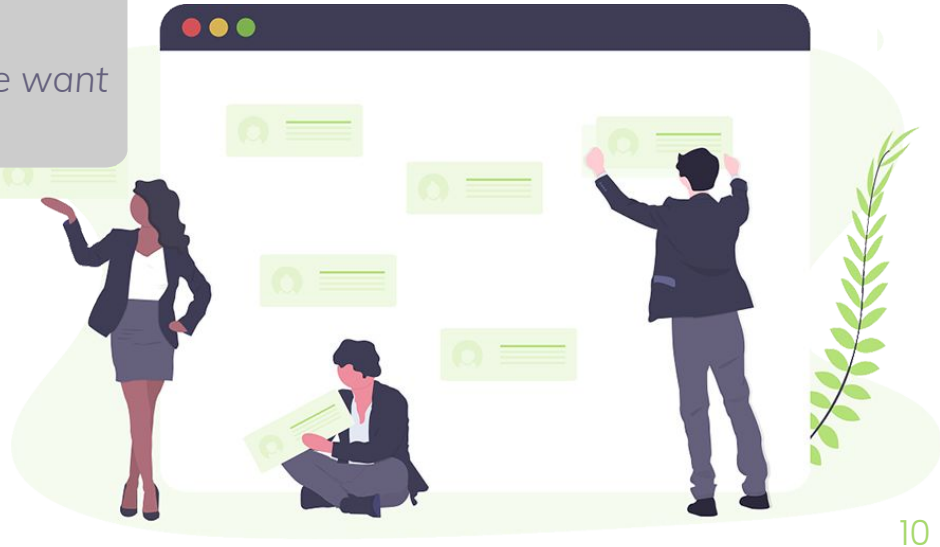
Positive or Negative?

Ignore Neutral, we only want strong opinions.

All 3 Sentiments?

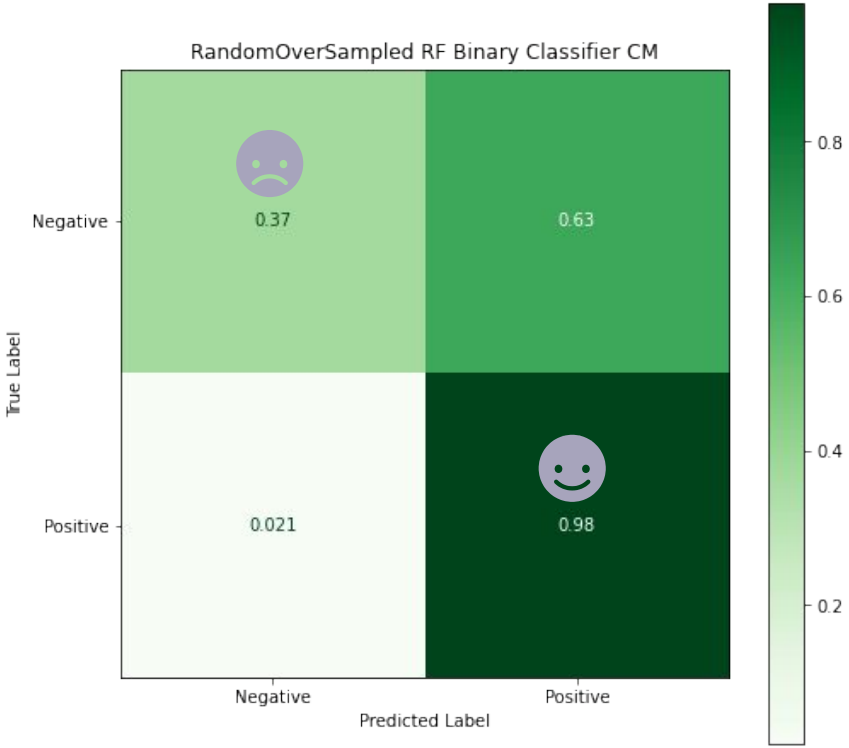
There were many Neutral tweets. We want them represented.

Create Models for Both!



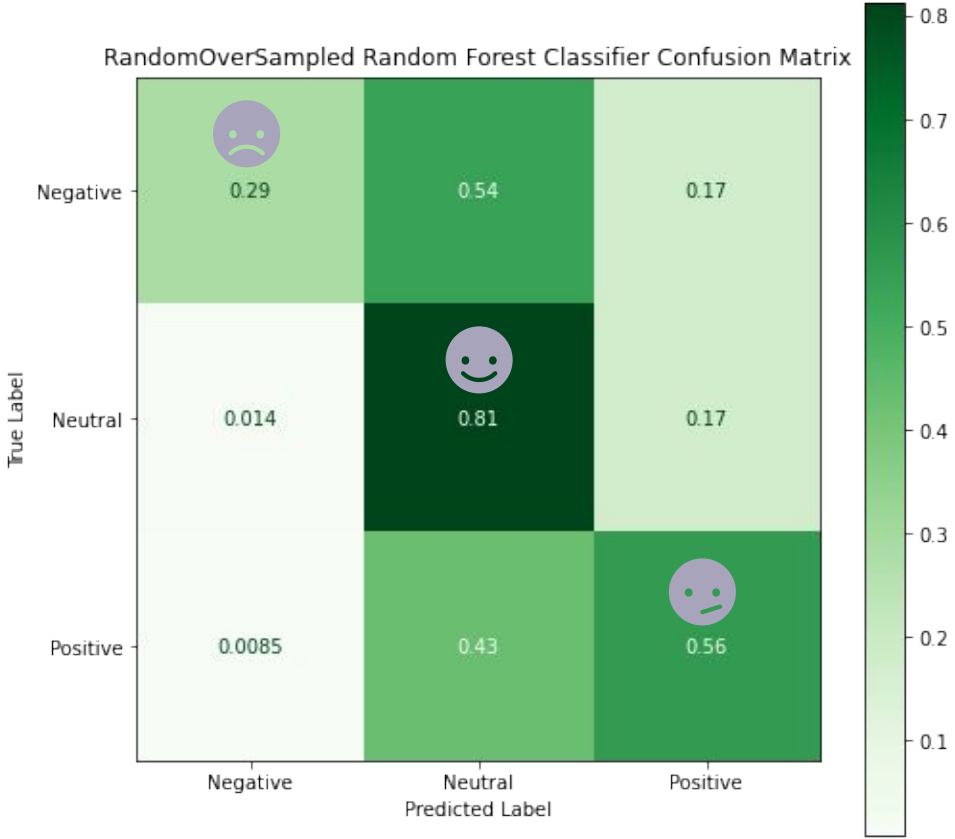
1 Pos I Neg Model Results

- Oversampling to address the issue of too few Negative tweets - not very successful
- Weighted F1 Score as Evaluation Metric for the (4) built models:
  - Useful for class imbalance
  - Balances Precision and Recall
- RandomForest w/ Oversampling generated the "best" model
  - Fantastic for Positive tweets
  - Poor for Negative tweets

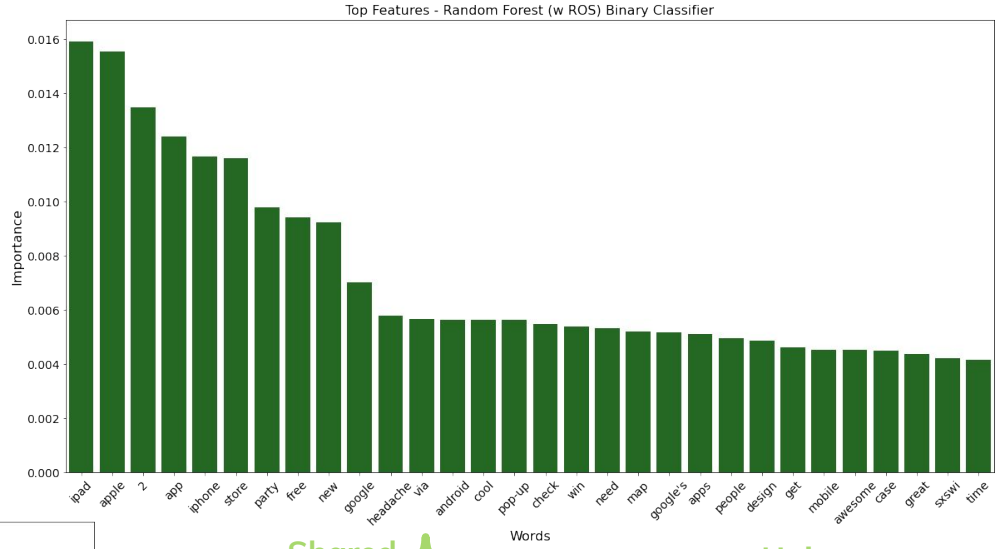


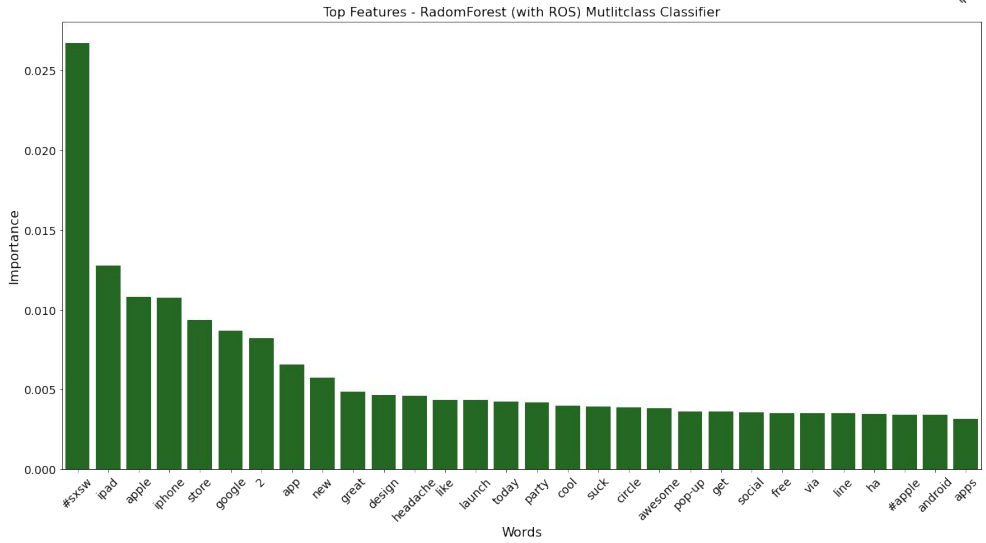
2 Pos /Neg /Neu Model Results

- Used Oversampling & Weighted F1 Score as Evaluation Metric for the (4) built models
- (RandomForest w/ Oversampling generated the “best” model
  - Good for Neutral tweets
  - Mediocre for Positive tweets
  - Poor for Negative tweets



What words mattered?





Shared

Apple	Party
iPad	New
iPhone	Cool
Google	Great
App	Free
Design	Headache

Unique

- Ha
- Circle
- Case
- Need
- Line
- Time

Future Work

1. Punctuation as a Feature?
2. SMOTE for class imbalance
3. Hyper-parameter tuning
4. Better understanding of LIME
5. Transfer Learning with GloVe
6. More Tweets!
    a. VADER-generate sentiment?

Thanks!

Any questions?

You can find me at:

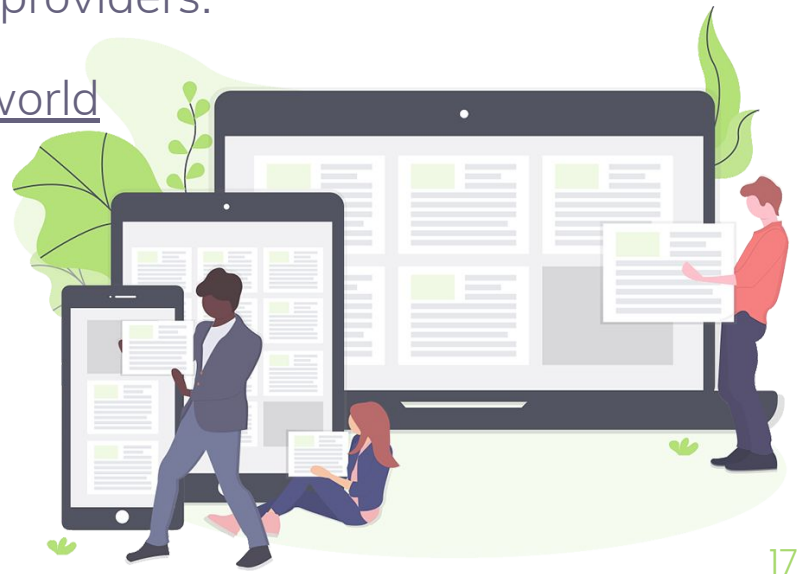
- https://github.com/lspope/
- https://www.linkedin.com/in/leahspope/
- leah@metisconsultingllc.com

Appendix

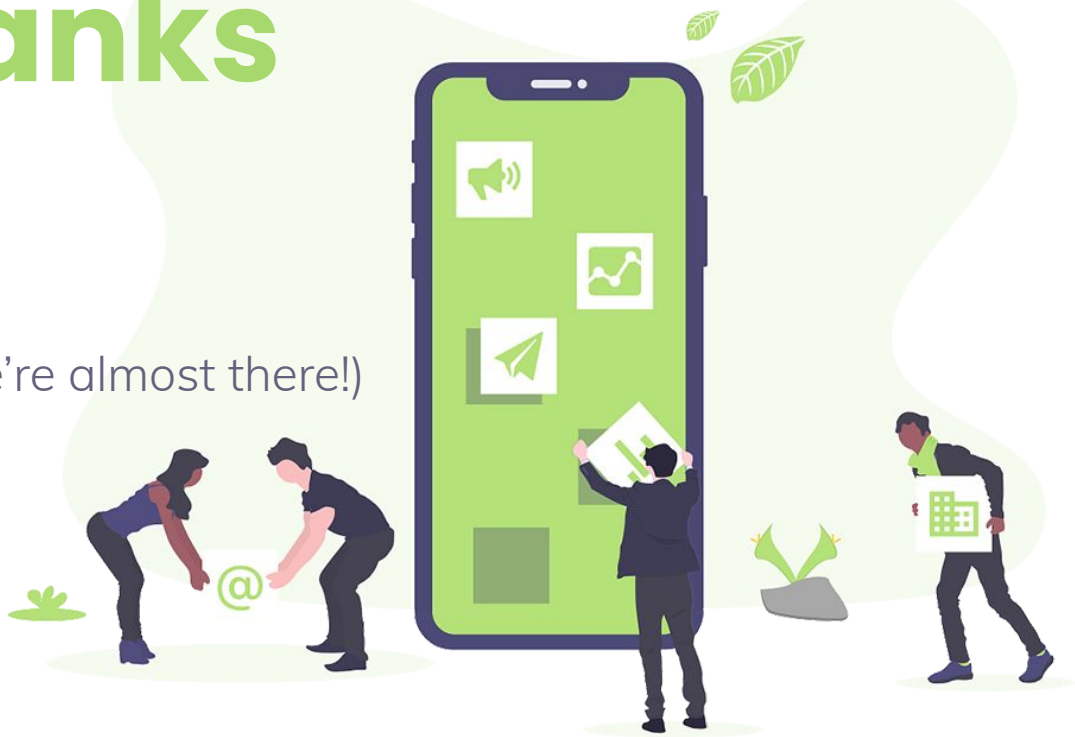
Credits

Special thanks to (awesome & free) resources providers:

- Data provided by CrowdFlower via data.world
- Presentation template by SlidesCarnival
- Photographs by Unsplash
- Illustrations by Undraw.co



Special Thanks

To special people:

- Family & Friends
- Fellow Cohort members (We're almost there!)
- Flatiron Instructors