

# What types of news content do Facebook users engage with most?

Using data from The New York Times

By Jessica Miles  
July 17, 2021

# Agenda

- Challenge statement: What and why
- My approach
- Dataset overview
- Machine learning results
- Recommendations to drive higher engagement

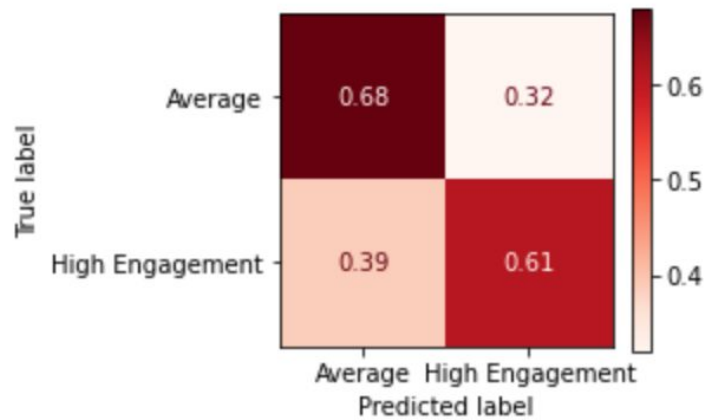
# The challenge: Posting content to engage Facebook users

- Facebook's News Feed algorithm prioritizes posts with more engagement (among other criteria)
- Better News Feed prioritization may reach more people and lead to new subscribers for the Times
- Not all available content can be posted to Facebook
- Multiple formats of the same content may exist (text, slideshow, video, audio, etc.)



# My approach - Machine Learning for Inference

- Trained simple machine learning models on post and article text, and metadata
  - Models included Naive Bayes, Logistic Regression
  - Tested binary and multi-class; binary performed best
- Best models identified about 61% of high engagement posts correctly
- Compared predictors of high and average engagement to interpret patterns



# Dataset Overview

# Dataset - All NYT Facebook Posts from Nov 2012 to Nov 2016

## All Facebook Posts - 48,000

- Features modeled:
  - Post text, post type, timing of post



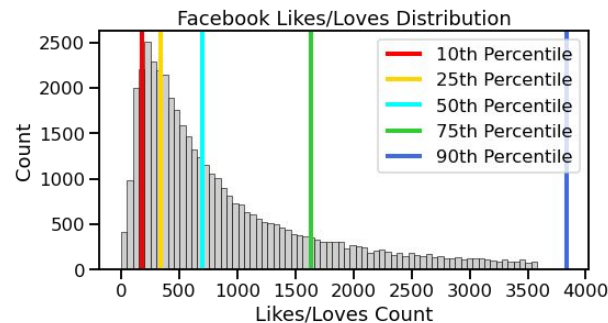
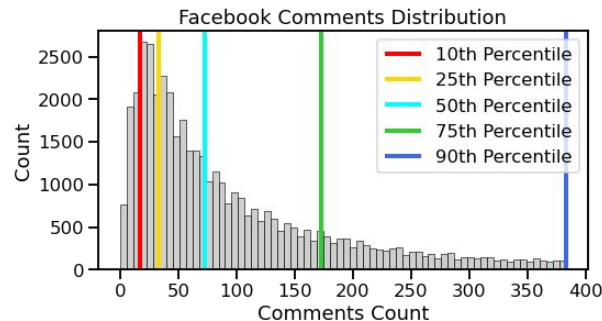
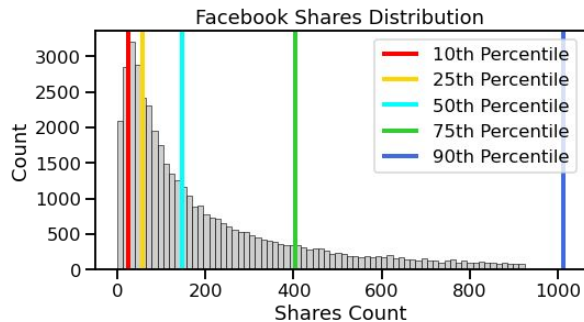
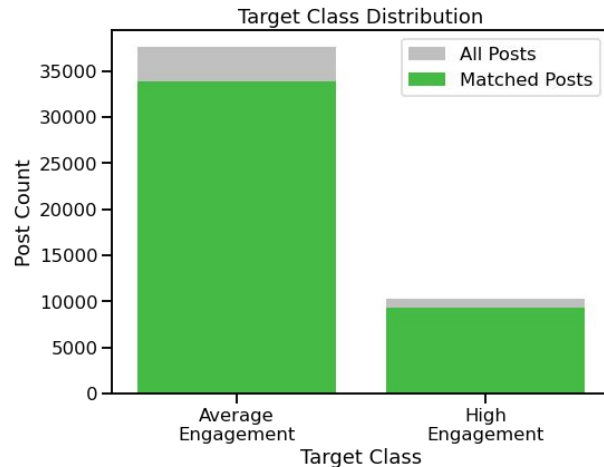
## Posts that could be matched to NYT articles from API - 43,000

- Features modeled:
  - Abstract and headline, news section, topical subjects, article word count, content type
- Not all posts could be matched, so modeled as a separate dataset

# Engagement Distribution

- Engagement metric distributions had many outliers
- Calculated percentile in each metric, then took the mean of the percentiles
- High engagement = Mean percentile > 75**

- Class distribution for all posts versus matched posts was comparable



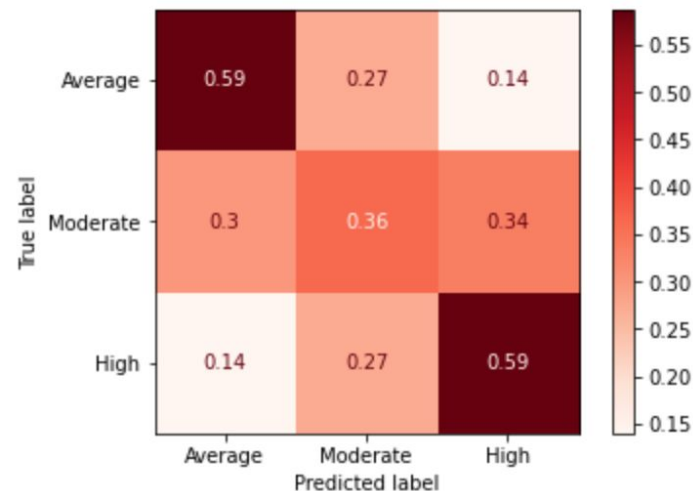
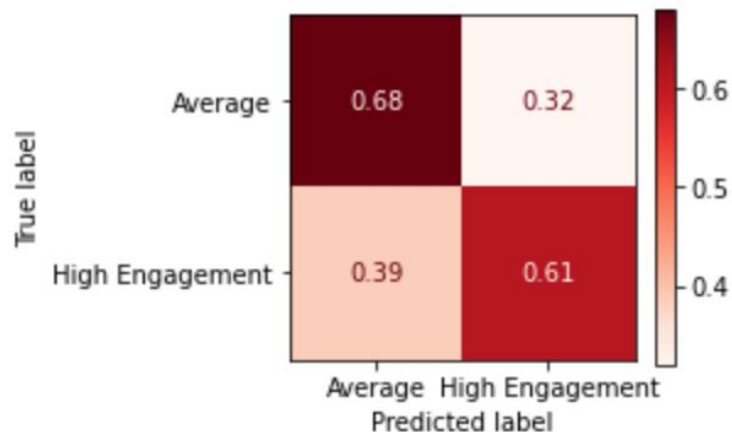




# Machine Learning Results

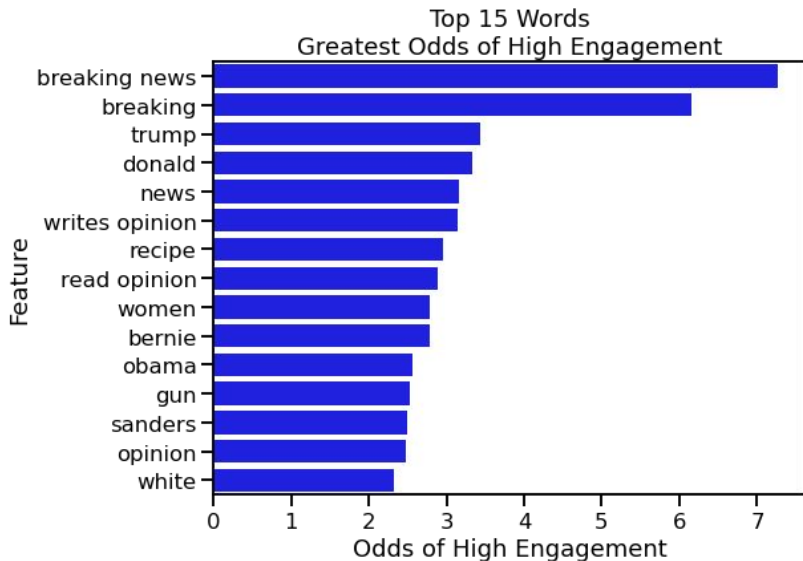
# Model Performance

- Binary model (75th percentile cutoff) performed best
- Multi-class model created new class for < 25th percentile
  - Performed about as well on High Engagement as binary
  - Moderate Engagement not predicted very well
- Engagement level tapers smoothly; no clear cutoff point for High versus Average. Some confusion is natural.



# Top Predictors of High Engagement

## Words in Facebook Posts



## Topics

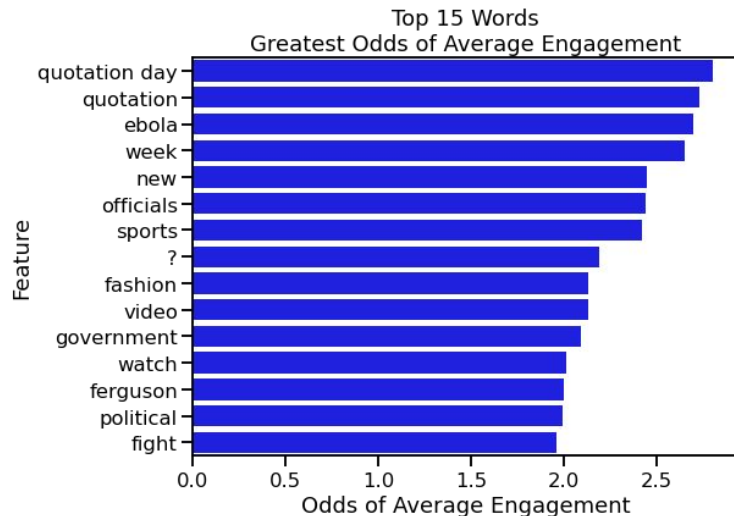
Feature	Odds of High Engagement
Archaeology And Anthropology	2.32
Presidential Election Of 2016	2.16
Cats	2.04
Walking	1.93
Happiness	1.89
Olympic Games (2016)	1.88
Deaths (Obituaries)	1.88
Muslim Americans	1.84
Chronic Traumatic Encephalopathy	1.84
Hair	1.83
Iphone	1.81
Beauty Contests	1.81
First Ladies (Us)	1.79
Animal Abuse, Rights And Welfare	1.77
Mathematics	1.77

## Post Characteristics

Feature	Odds of High Engagement
Post Type - Video	1.77
Post Type - Photo	1.52

Feature	Odds of High Engagement
Post Time - 7 PM to 11 PM	1.28
Post Time - Weekend Day	1.26

## Words in Facebook Posts\*\*



## News Sections

Feature	Odds of Average Engagement
Technology	1.99
Books	1.77
New York	1.73
Blogs	1.67

Feature	Odds of Average Engagement
Multimedia/Photos	1.65
Education	1.63
World	1.6
Sports	1.54

# Top Predictors of Average Engagement

## Topics

Feature	Odds of Average Engagement
Midterm Elections (2014)	2.69
Mobile Applications	2.16
Automobile Safety Features And Defects	2.16
Ebola Virus	2
Murders And Attempted Murders	1.99
Middle East And North Africa Unrest	1.88
Politics And Government	1.87
Political Advertising	1.82
Tests (Medical)	1.8
Athletics And Sports	1.79
E-Commerce	1.78
Jury System	1.74
Brussels Attacks (March 2016)	1.74
Alternative And Renewable Energy	1.73
Espionage And Intelligence Services	1.69

\*\* Exact odds of word predictors varied depending on test split  
Recommendations are based on analysis of median engagement as well as model odds

# Recommendations to Increase Engagement

# What to prioritize?

## Popular Topics:

- Breaking News
- Obituaries
- Recipes
- Current president and Presidential election
- Controversial and polarizing topics:
  - Opinion pieces
  - Gun violence, Gender and racial equality, Vaccination
- "Fun" topics:
  - Examples are Cats, Happiness, Hair, Kids

## How and When to Post:

- In the evening (7-11 PM Eastern)
- On weekends
- As video or photo uploads instead of links, when appropriate

# What to de-prioritize?

## Less popular Topics:

- Regularly recurring features:
  - "Quotation of the Day", "New York Today", "Daily Briefing", "What You Should Watch"
- General government and politics
- General sports coverage
- Fashion, Technology, Books, Blogs
- World coverage

## Words to Avoid:

- "What you need to know"
- "week"

# Caveats

- Facebook's own News Feed algorithm is very important to driving engagement, and is based partly on user-centric preferences which we can't model
- Odds of top predictors varied slightly depending on data split used for training
- Leaning in to controversial or divisive topics should be carefully considered
- The cutoff point for "High engagement" is somewhat arbitrary
- Tastes change, so results from 2016 may not be applicable to present day



# Potential Future Work

- Review sentiment of articles to see whether that affects engagement
- Compare engagement on Facebook to comments count on the New York Times website, so see if there is a difference in what drives engagement there
- Create an interactive dashboard so engagement of certain words and subjects can be reviewed

# Thank you for reading!

For questions or comments, please  
contact:

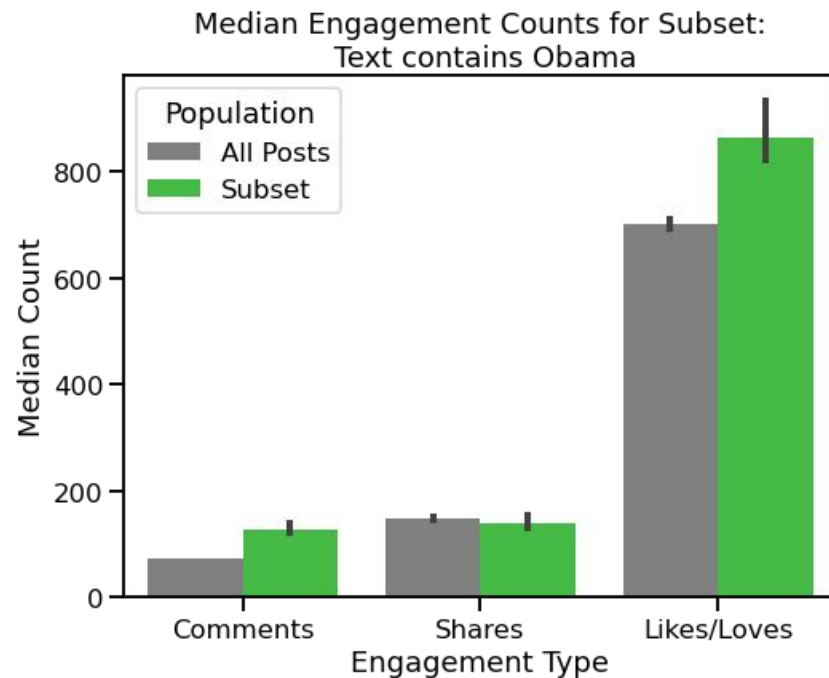
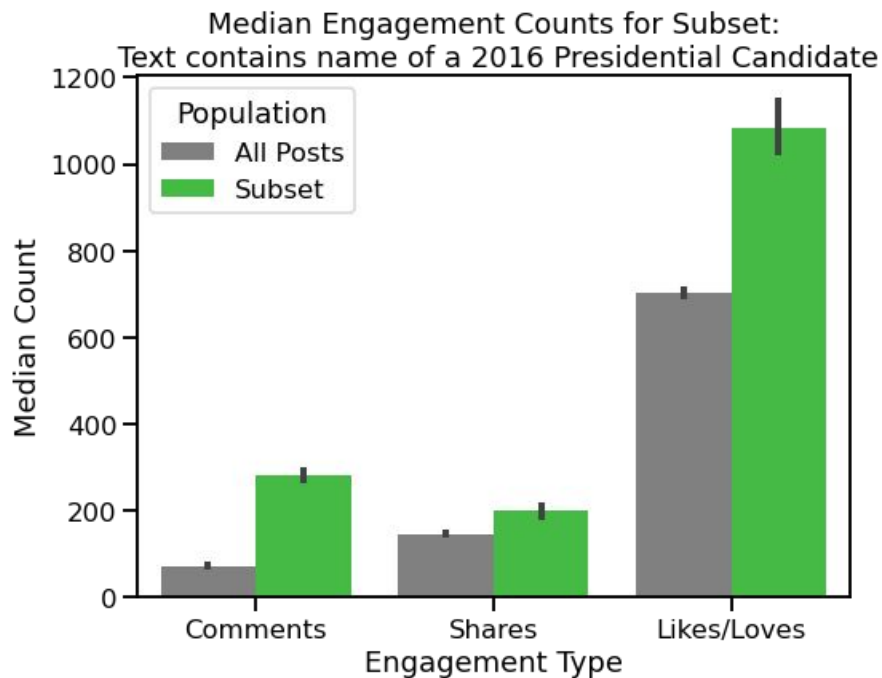
Jessica Miles

[jess.c.miles@gmail.com](mailto:jess.c.miles@gmail.com)

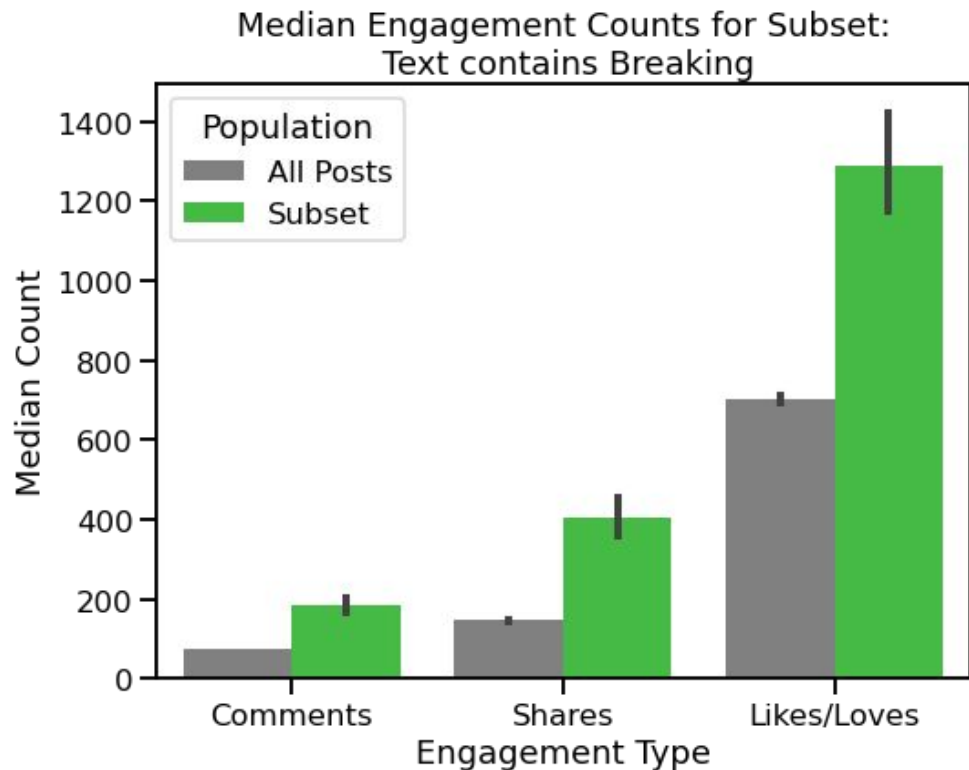
— — —

# Appendix - Engagement per Significant Feature

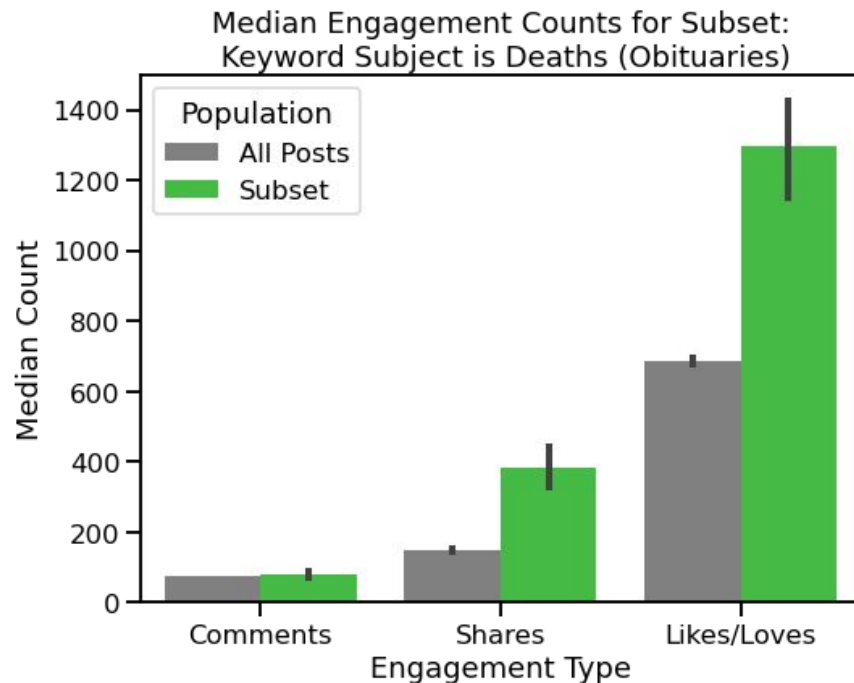
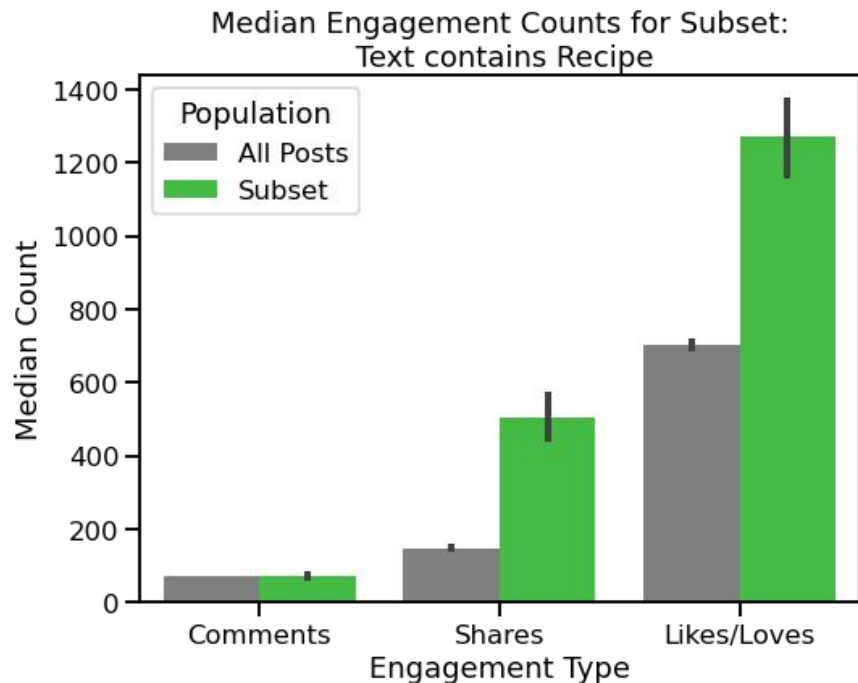
# More Engaging - Presidential Topics



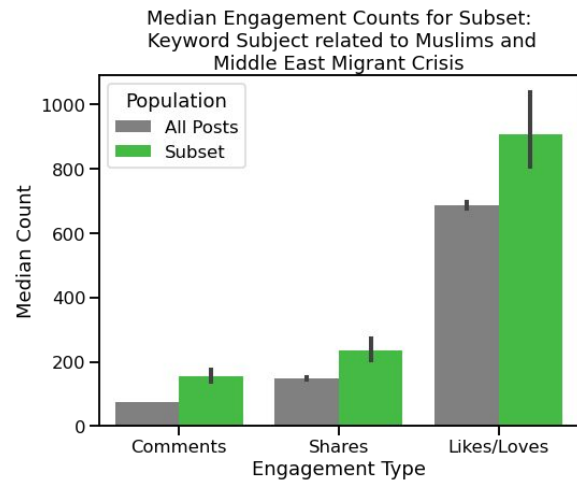
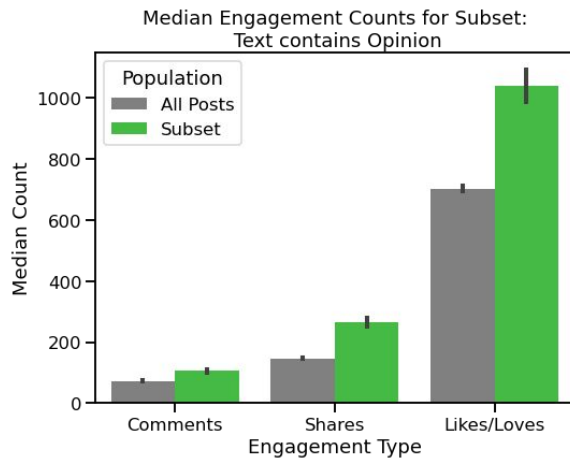
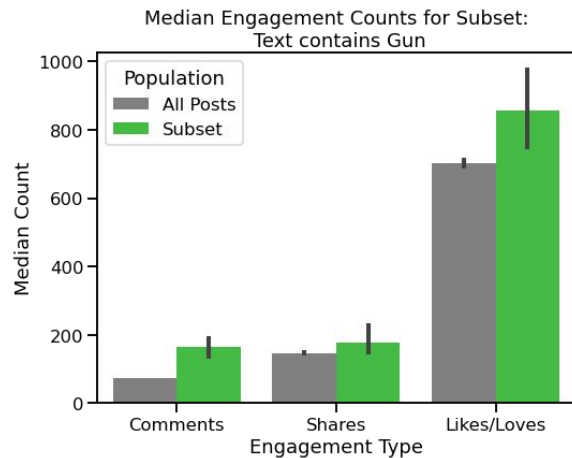
# More Engaging - Breaking News



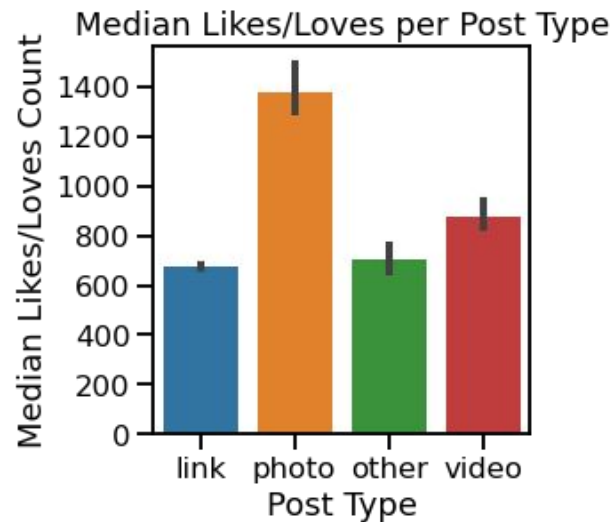
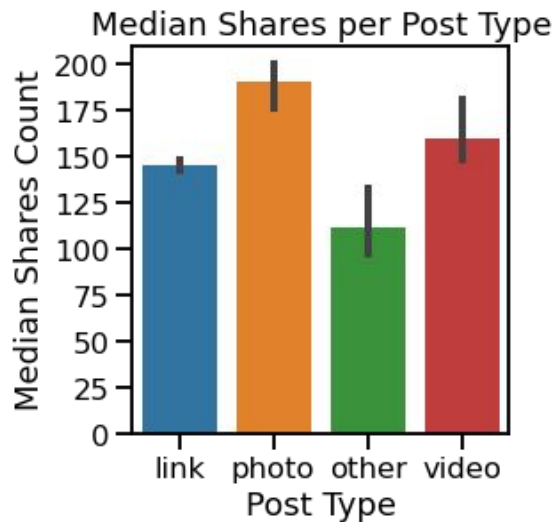
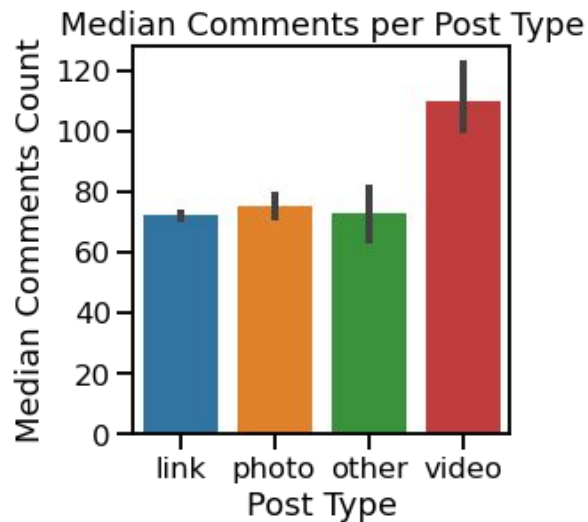
# More Engaging - Recipes and Obituaries



# More Engaging - Controversial Topics



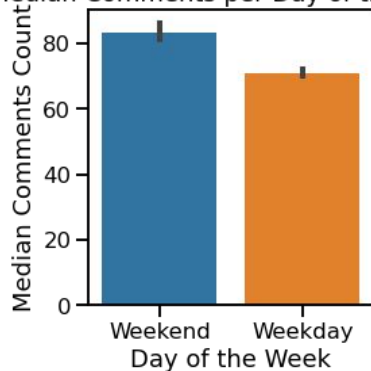
# More Engaging - Post Type



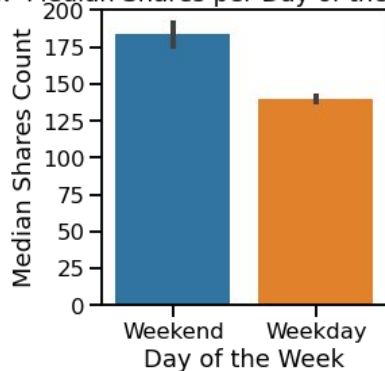


# More Engaging - Post Timing

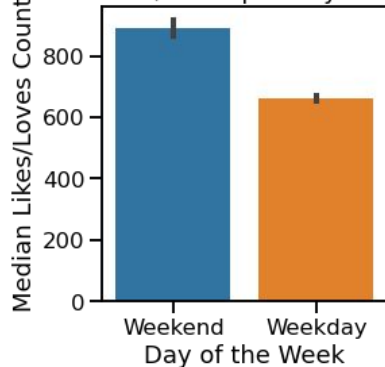
Median Comments per Day of the Week



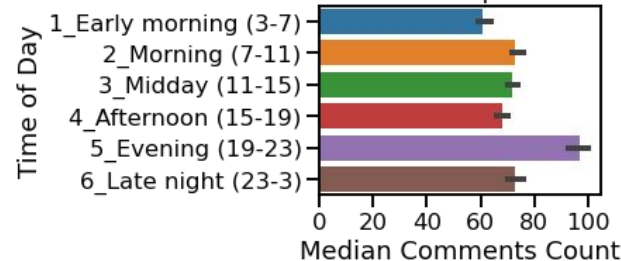
Median Shares per Day of the Week



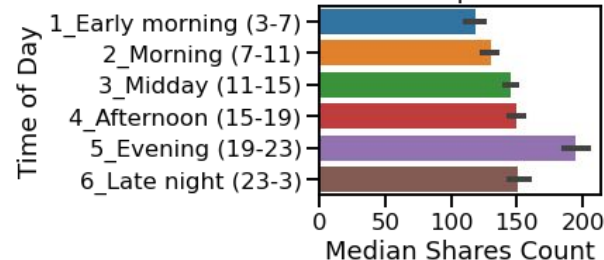
Median Likes/Loves per Day of the Week



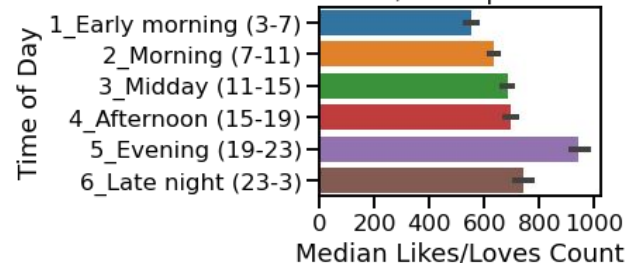
Median Comments per Time of Day



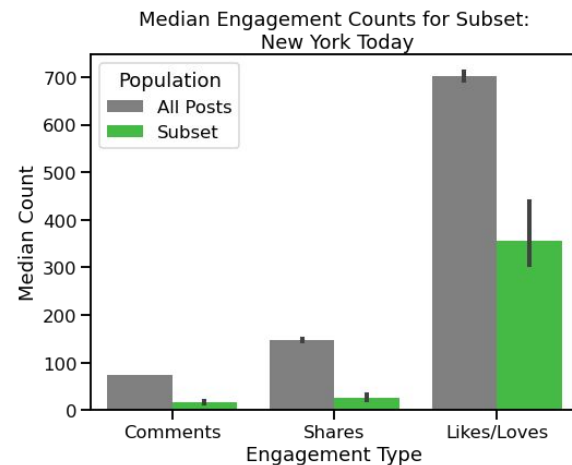
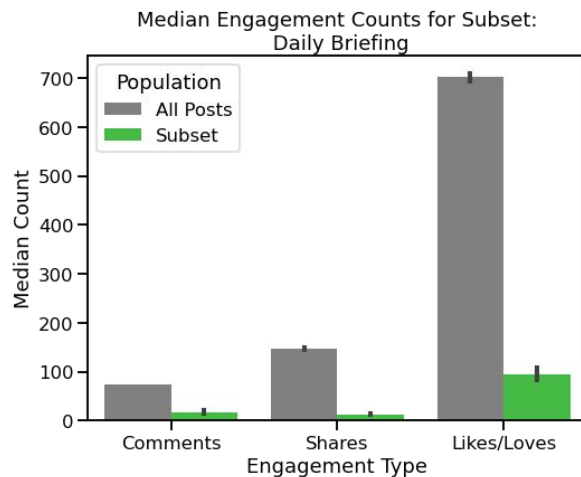
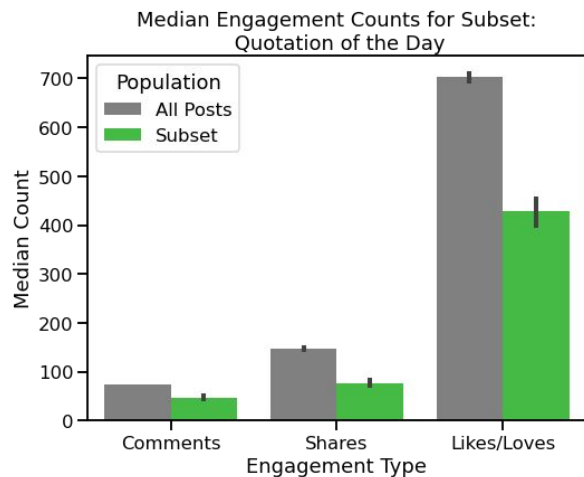
Median Shares per Time of Day



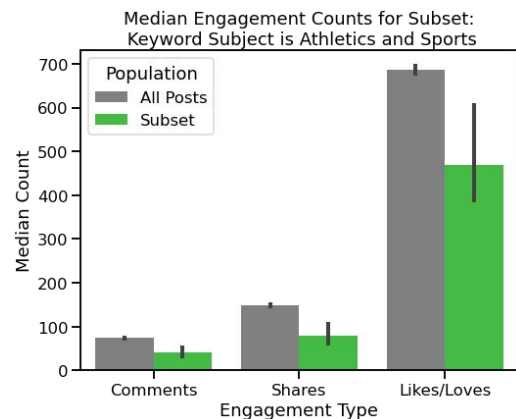
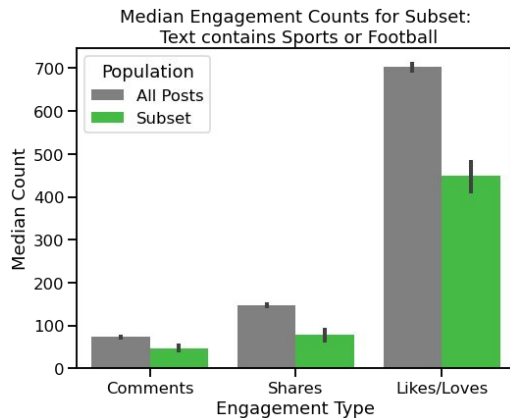
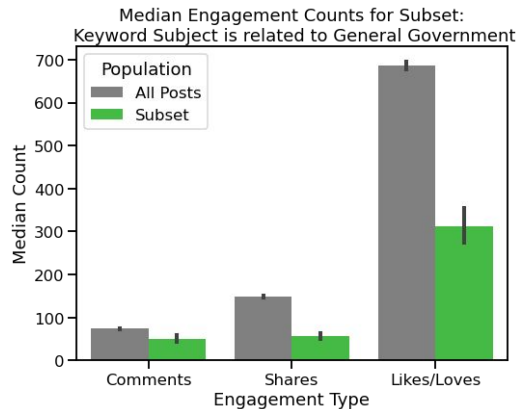
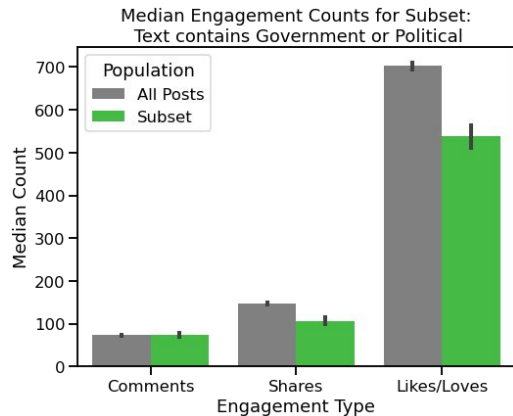
Median Likes/Loves per Time of Day



# Less Engaging - Recurring Features



# Less Engaging - General Sports



# Less Engaging - Certain Keywords

