Project Check-In

By: Grant, Wenxuan Gu, Seok Jun Seo, Michael Huang, & Sam Burch (Group 8)

Most Influential Predictors

- Many significant predictors (20+)

  - Because of the similarity, we can

    reduce this number

- Potentially most important predictors

  - Key Word Average

  - Number of words in Article

  - Global Subjectivity

```
                               Pr(>|t|)
(Intercept)                    0.214542
n_tokens_title                 0.000426 ***
n_tokens_content               0.000368 ***
n_non_stop_words               0.000180 ***
num_hrefs                      5.18e-05 ***
num_imgs                       1.87e-08 ***
average_token_length           2.89e-09 ***
kw_max_min                     0.047487 *
kw_avg_min                     0.063472 .
kw_avg_max                     0.043976 *
kw_min_avg                     2.78e-05 ***
kw_max_avg                     2.21e-05 ***
kw_avg_avg                     5.52e-09 ***
weekday_is_tuesday             0.123639
weekday_is_wednesday           0.000568 ***
weekday_is_friday              0.062483 .
LDA_02                         0.001008 **
LDA_03                         0.046755 *
global_subjectivity            0.000760 ***
global_sentiment_polarity      0.060525 .
rate_positive_words            0.105487
max_negative_polarity          0.017637 *
title_sentiment_polarity       0.036504 *
abs_title_sentiment_polarity   0.144252
```

Weekdays and Weekends

Average Shares by Weekday
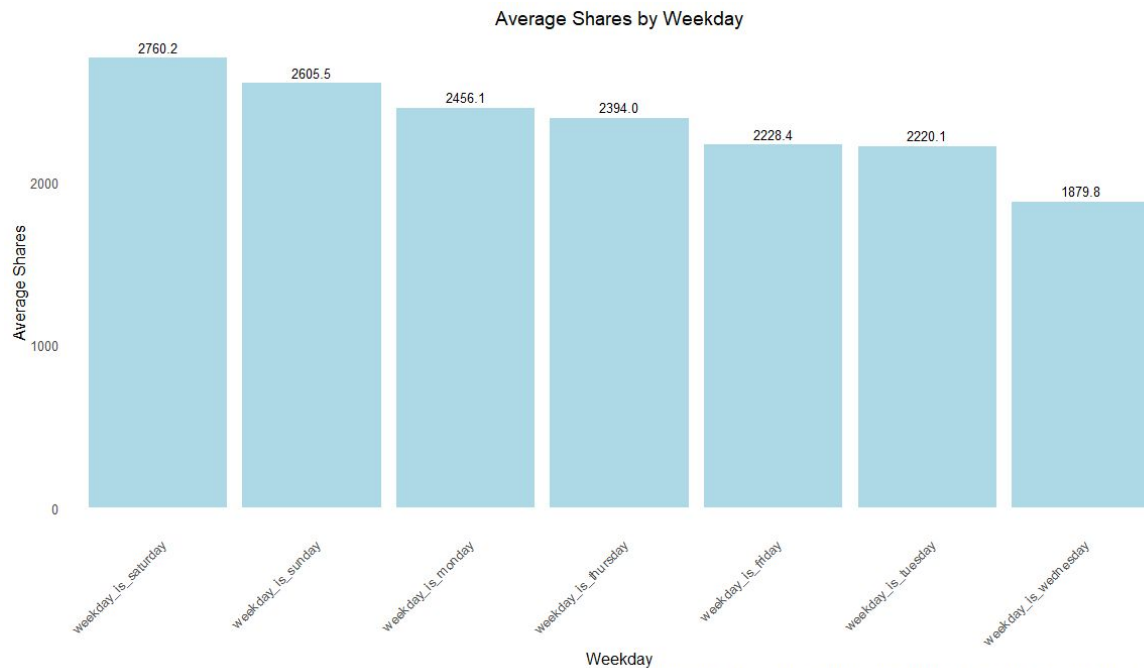
Source: https://archive.ics.uci.edu/dataset/332/online+news+popularity. Sample size: n = 8427.
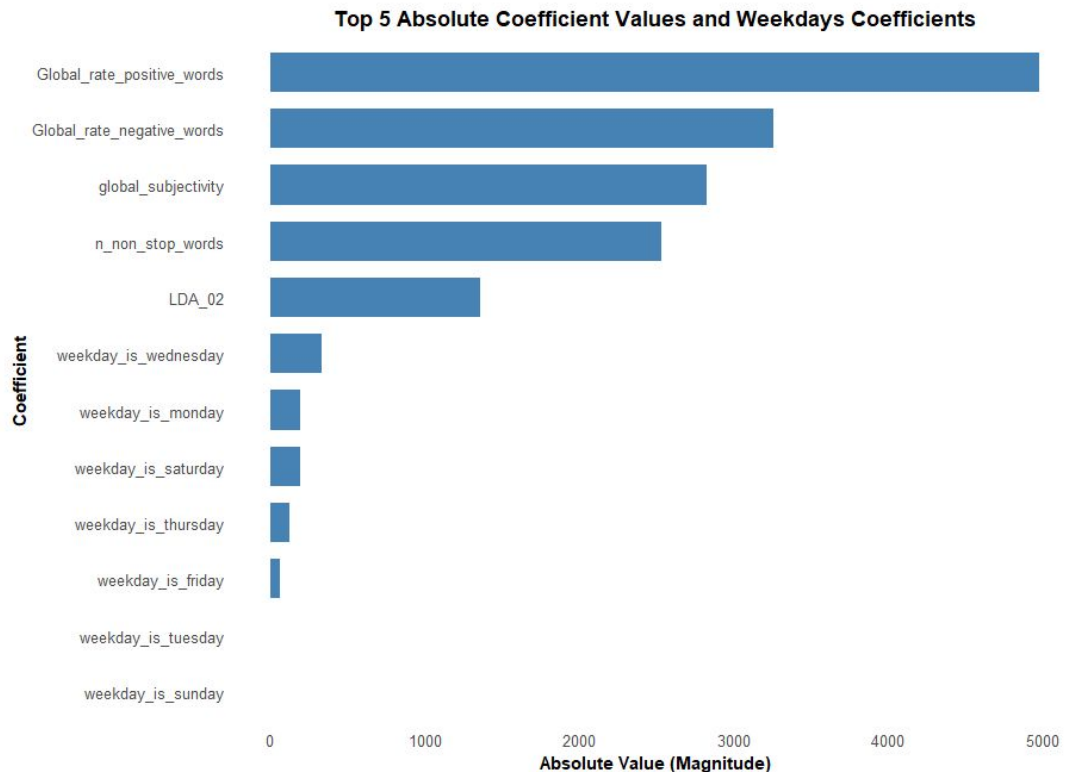
Weekdays and Weekends

- Wednesday has the lowest number of average shares

- Saturday has the highest number of average shares

- Difference between the average number of each weekday is less than 1000

- Mean of average shares: 2363.44

- Suggesting weekday_is_* has low influence on number of shares

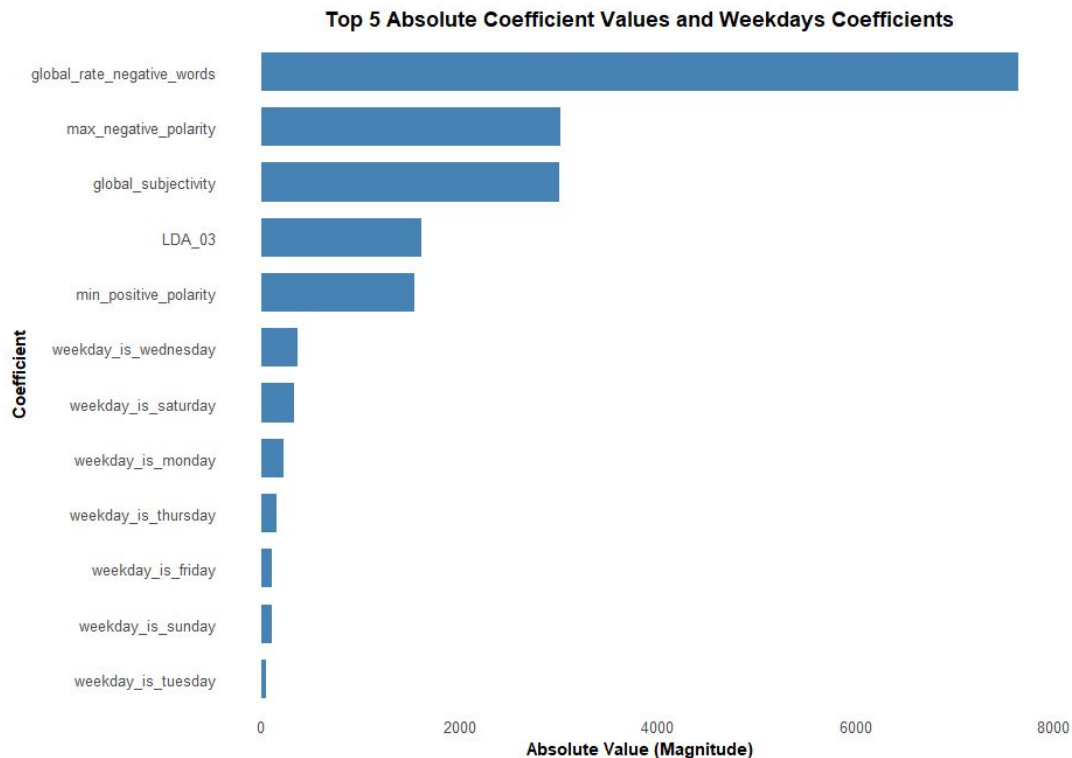- Suggesting Wednesday will result in a lower number of shares

Machine Learning Approach(Lasso)

Top 5 Absolute Coefficient Values and Weekdays Coefficients

Source: https://archive.ics.uci.edu/dataset/332/online+news+popularity. Sample size: 8427

Machine Learning Approach(Ridge)

Top 5 Absolute Coefficient Values and Weekdays Coefficients

Source: https://archive.ics.uci.edu/dataset/332/online+news+popularity. Sample size: 8427

Polarity – Positive Word Rate/Title Subjectivity

1	Neutral Avg	Mod Pos Avg	Very Pos Avg	Mod Neg Avg	Very Neg Avg
2	2018	2287	2890	1585	3458

Table shows the Avg. shares per group based on the rate of positive words that are found in the article.

1	Neutral Title	Mod Sub Avg	Very Sub Avg	Mod Obj Avg	Very Obj Avg
2	2532	2633	2446	2150	2116

Table shows the Avg. shares per group based on the subjectivity of the title in the articles.
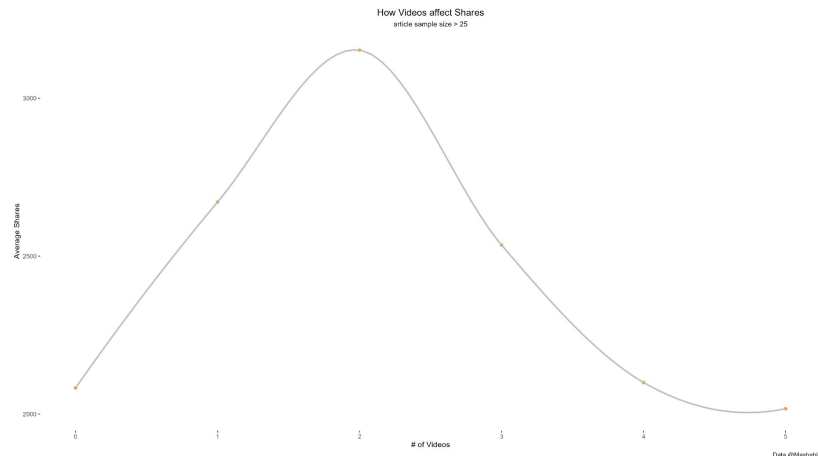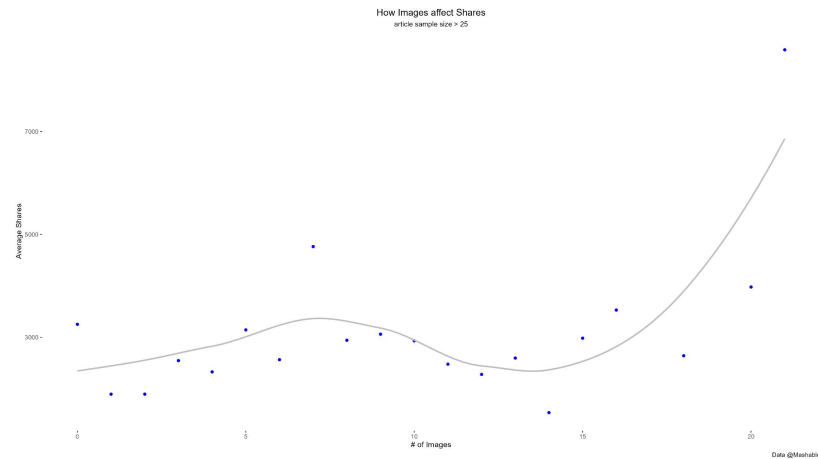
Polarity Findings

- Both extremes, very high and very low positive word rates perform the best when it comes to the number of shares

- A title with a neutral to a very subjective title outperforms those with an objective title.

Images vs Videos

- Images seem slightly more important than Videos

- (Possible) target images amounts: 8-10 & 15-20

- A couple videos are better than none

- There is no interaction between videos and images

How Images affect Shares
article sample size > 25

Average Shares

# of Images

Data @Mashable

How Videos affect Shares
article sample size > 25

Average Shares

# of Videos

Data @Mashable

Our Questions

- World articles

- Variable Names

    - keywords

    - LDA

    - polarity

- Exact model vs. what's the most predictive

- Largest coefficient or Most significant