online-news-popularity

Sneh Gurdasani, Akshit Jain, Farhan Ansari, Sagar Singh

11/11/2019

Load and Tidy Dataset

```
df <- read_csv("OnlineNewsPopularity/OnlineNewsPopularity.csv", col_types = cols())
get_tidy_data(df)
```

```
## # A tibble: 6 x 61
##    url   timedelta n_tokens_title n_tokens_content n_unique_tokens
##    <chr>     <dbl>          <dbl>            <dbl>           <dbl>
## 1 http~       731             12              219           0.664
## 2 http~       731              9              255           0.605
## 3 http~       731              9              211           0.575
## 4 http~       731              9              531           0.504
## 5 http~       731             13             1072           0.416
## 6 http~       731             10              370           0.560
## # ... with 56 more variables: n_non_stop_words <dbl>,
## #   n_non_stop_unique_tokens <dbl>, num_hrefs <dbl>, num_self_hrefs <dbl>,
## #   num_imgs <dbl>, num_videos <dbl>, average_token_length <dbl>,
## #   num_keywords <dbl>, data_channel_is_lifestyle <dbl>,
## #   data_channel_is_entertainment <dbl>, data_channel_is_bus <dbl>,
## #   data_channel_is_socmed <dbl>, data_channel_is_tech <dbl>,
## #   data_channel_is_world <dbl>, kw_min_min <dbl>, kw_max_min <dbl>,
## #   kw_avg_min <dbl>, kw_min_max <dbl>, kw_max_max <dbl>,
## #   kw_avg_max <dbl>, kw_min_avg <dbl>, kw_max_avg <dbl>,
## #   kw_avg_avg <dbl>, self_reference_min_shares <dbl>,
## #   self_reference_max_shares <dbl>, self_reference_avg_sharess <dbl>,
## #   weekday_is_monday <dbl>, weekday_is_tuesday <dbl>,
## #   weekday_is_wednesday <dbl>, weekday_is_thursday <dbl>,
## #   weekday_is_friday <dbl>, weekday_is_saturday <dbl>,
## #   weekday_is_sunday <dbl>, is_weekend <dbl>, LDA_00 <dbl>, LDA_01 <dbl>,
## #   LDA_02 <dbl>, LDA_03 <dbl>, LDA_04 <dbl>, global_subjectivity <dbl>,
## #   global_sentiment_polarity <dbl>, global_rate_positive_words <dbl>,
## #   global_rate_negative_words <dbl>, rate_positive_words <dbl>,
## #   rate_negative_words <dbl>, avg_positive_polarity <dbl>,
## #   min_positive_polarity <dbl>, max_positive_polarity <dbl>,
## #   avg_negative_polarity <dbl>, min_negative_polarity <dbl>,
## #   max_negative_polarity <dbl>, title_subjectivity <dbl>,
## #   title_sentiment_polarity <dbl>, abs_title_subjectivity <dbl>,
## #   abs_title_sentiment_polarity <dbl>, shares <dbl>
```

```
get_tidy_data <- function(df) {
  head(df)
}
```