

# Final Project: Predicting Article Popularity in Online News Media Using Machine Learning

Xiangyu Xiao and Binyu Yang

2024-04-11

## 1. Data Preparation

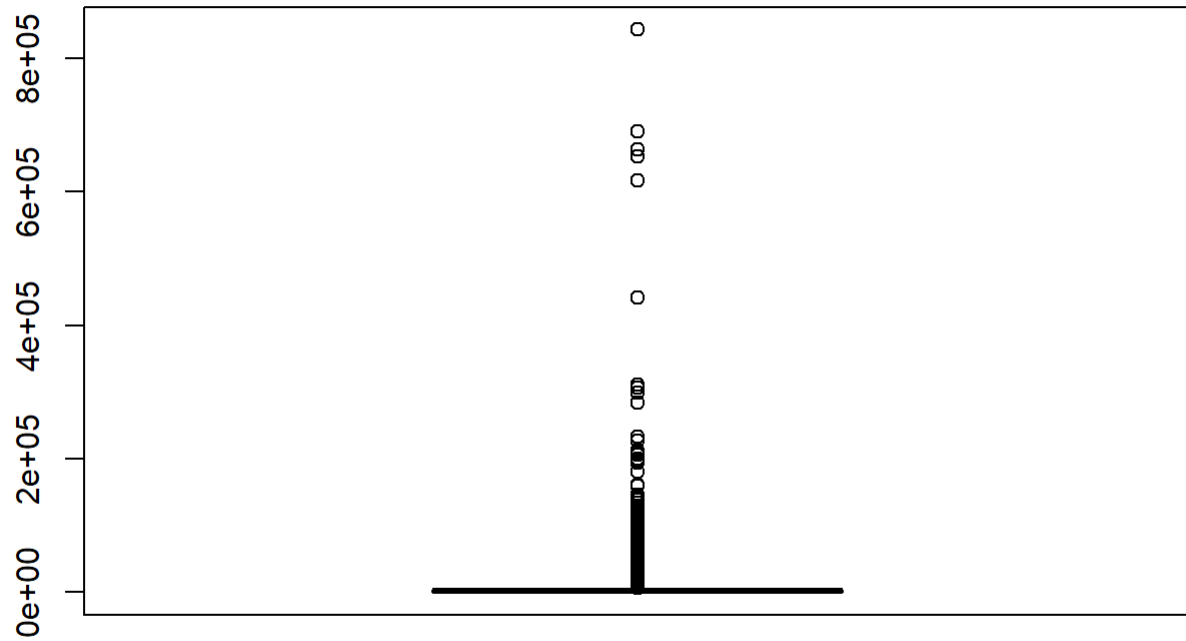
```
# Load the dataset
data <- read.csv("D:/MBAN/Machine Learning and AI/FInal/online+news+popularity/OnlineNewsPopularity/OnlineNewsPopularity.csv")
```

```
# Check for missing values
sum(is.na(data))
```

```
## [1] 0
```

```
# There are no missing values in the dataset
```

```
# Check for outliers of the target variable
boxplot(data$shares)
```



*# As we see many outliers here, we need to remove those outliers. We use the IQR method to remove the outliers.*

```
# Remove outliers
library(dplyr)
```

```
##
## 载入程辑包: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
# Calculate the quartiles and IQR for the variable
Q1 <- quantile(data$shares, 0.25)
Q3 <- quantile(data$shares, 0.75)
IQR <- Q3 - Q1

# Define the bounds for what is considered an outlier
lower_bound <- Q1 - 1.5 * IQR
upper_bound <- Q3 + 1.5 * IQR

# Filter out the rows where the variable is an outlier
data_no_outliers <- data %>%
  filter(data$shares >= lower_bound & data$shares <= upper_bound)
```

## 2. EDA

```
# Correlation between the variables
# Select only numeric columns
numeric_data <- select_if(data_no_outliers, is.numeric)

# Now you can compute the correlation matrix
correlation_matrix <- cor(numeric_data)

# View the correlation matrix
print(correlation_matrix)
```

##		timedelta	n_tokens_title	n_tokens_content
## timedelta		1.0000000000	-2.360087e-01	-0.0727660646
## n_tokens_title		-0.2360086927	1.000000e+00	0.0315690449
## n_tokens_content		-0.0727660646	3.156904e-02	1.0000000000
## n_unique_tokens		0.1978728044	-5.604320e-02	-0.4165977942
## n_non_stop_words		0.1464907153	-3.987952e-02	0.1992112091
## n_non_stop_unique_tokens		0.1782751349	-4.812465e-02	-0.2385936226
## num_hrefs		0.0004534453	-4.207178e-02	0.4269887229
## num_self_hrefs		0.0682016912	-6.599999e-03	0.3036560429
## num_imgs		-0.0205447979	5.176119e-04	0.3570744479
## num_videos		-0.0055416020	5.273517e-02	0.1071367029
## average_token_length		0.1192984922	-6.779093e-02	0.1642624416
## num_keywords		0.0559806387	-1.522163e-03	0.0706915910
## data_channel_is_lifestyle		0.0567933203	-7.058041e-02	0.0322860346
## data_channel_is_entertainment		-0.0495322721	1.346361e-01	0.0649442760
## data_channel_is_bus		0.0625109118	-2.855529e-02	-0.0182403977
## data_channel_is_socmed		0.0730013748	-8.774792e-02	0.0284352527
## data_channel_is_tech		0.0831765593	-4.398069e-02	0.0175663823
## data_channel_is_world		-0.1770178793	4.649488e-02	0.0601794864
## kw_min_min		0.5917295555	-1.062407e-01	-0.0573752026
## kw_max_min		0.0324452683	-4.728459e-03	0.0027778389
## kw_avg_min		0.1475609966	-3.400937e-02	-0.0019092958
## kw_min_max		-0.0770000797	1.057968e-02	-0.0223714013
## kw_max_max		-0.6399255709	1.184967e-01	0.0636262772
## kw_avg_max		-0.4945883931	1.052795e-01	-0.0902590993
## kw_min_avg		-0.1572421960	-2.049816e-03	-0.0271703033
## kw_max_avg		-0.0518910395	6.275237e-03	-0.0278774252
## kw_avg_avg		-0.1640704979	2.881137e-03	-0.0789510010
## self_reference_min_shares		-0.0133254851	9.422205e-05	-0.0267235773
## self_reference_max_shares		-0.0156088242	1.031335e-03	0.0266344293
## self_reference_avg_share		-0.0165543059	3.141392e-03	-0.0112332741
## weekday_is_monday		-0.0035072945	2.381394e-03	0.0004503348
## weekday_is_tuesday		-0.0029793827	7.871474e-03	-0.0051868052
## weekday_is_wednesday		0.0086005550	1.062032e-02	-0.0214339423
## weekday_is_thursday		0.0027366392	-1.417761e-02	-0.0060066282
## weekday_is_friday		-0.0034942650	-1.737829e-03	-0.0145213416
## weekday_is_saturday		-0.0102980693	-1.210884e-02	0.0382053642
## weekday_is_sunday		0.0068943950	3.532759e-03	0.0345137883
## is_weekend		-0.0021973737	-6.010391e-03	0.0531709960
## LDA_00		0.0827905949	-7.119752e-02	0.0113384246
## LDA_01		0.0018283017	6.334422e-02	-0.0029827251
## LDA_02		-0.1492858889	3.662239e-02	0.0918593634
## LDA_03		-0.0235064895	4.305753e-02	-0.1309689876
## LDA_04		0.0952387917	-6.260302e-02	0.0304683515
## global_subjectivity		0.1267584647	-5.331063e-02	0.1301856420
## global_sentiment_polarity		0.1612219787	-6.798157e-02	0.0129594896
## global_rate_positive_words		0.2049534243	-5.989166e-02	0.1287556595
## global_rate_negative_words		0.0023847653	1.673360e-02	0.1323259212
## rate_positive_words		0.1960345318	-6.321368e-02	0.0863511045
## rate_negative_words		-0.0822885533	3.534200e-02	0.1059174453
## avg_positive_polarity		0.1212115431	-4.393267e-02	0.1330633428
## min_positive_polarity		0.0533339006	-2.512615e-02	-0.2652683292
## max_positive_polarity		0.0919924134	-1.371180e-02	0.4179651120
## avg_negative_polarity		0.0073441191	-1.850611e-02	-0.1374322819
## min_negative_polarity		0.0723377665	-3.471307e-02	-0.4567120829

## max_negative_polarity	-0.0619027531	1.430564e-02	0.2272498497
## title_subjectivity	-0.0174773106	8.498210e-02	0.0037789065
## title_sentiment_polarity	0.0399822493	5.280820e-03	0.0197520513
## abs_title_subjectivity	0.0088593215	-1.492605e-01	0.0109036178
## abs_title_sentiment_polarity	-0.0018525326	4.875231e-02	0.0114885660
## shares	0.0365717262	-4.204983e-02	0.0478207415
##	n_unique_tokens	n_non_stop_words	
## timedelta	0.1978728044	0.1464907153	
## n_tokens_title	-0.0560431954	-0.0398795194	
## n_tokens_content	-0.4165977942	0.1992112091	
## n_unique_tokens	1.0000000000	0.6671083247	
## n_non_stop_words	0.6671083247	1.0000000000	
## n_non_stop_unique_tokens	0.9371275719	0.7571318338	
## num_hrefs	-0.1214052016	0.1656508959	
## num_self_hrefs	-0.0596678473	0.1444486919	
## num_imgs	-0.2466785971	0.0121732380	
## num_videos	0.0198672457	0.0164966714	
## average_token_length	0.6435803470	0.9407069389	
## num_keywords	-0.0749778439	-0.0235368421	
## data_channel_is_lifestyle	-0.0104479842	0.0281317016	
## data_channel_is_entertainment	0.0012134419	-0.0007286326	
## data_channel_is_bus	0.0518389810	0.0647419326	
## data_channel_is_socmed	0.0133696646	0.0330464680	
## data_channel_is_tech	0.0044645151	0.0735953955	
## data_channel_is_world	-0.0773288219	-0.0005446489	
## kw_min_min	0.0913821222	0.0481146193	
## kw_max_min	-0.0017722446	0.0004074634	
## kw_avg_min	0.0139020688	0.0179855305	
## kw_min_max	-0.0052617458	-0.0403968837	
## kw_max_max	-0.0978279113	-0.0520226276	
## kw_avg_max	-0.0452925680	-0.1656019722	
## kw_min_avg	-0.0253407244	-0.0741840808	
## kw_max_avg	-0.0073578140	-0.0470906798	
## kw_avg_avg	-0.0299799112	-0.1422126841	
## self_reference_min_shares	0.0410505364	0.0319290374	
## self_reference_max_shares	0.0227748611	0.0408866857	
## self_reference_avg_shares	0.0423824535	0.0422485164	
## weekday_is_monday	0.0044950019	0.0056093043	
## weekday_is_tuesday	0.0032352228	0.0020998442	
## weekday_is_wednesday	0.0072990724	-0.0033731522	
## weekday_is_thursday	0.0034508201	0.0016293416	
## weekday_is_friday	0.0140612051	0.0006250605	
## weekday_is_saturday	-0.0380524274	-0.0092150652	
## weekday_is_sunday	-0.0121664317	-0.0010106533	
## is_weekend	-0.0363197025	-0.0073453165	
## LDA_00	0.0392700507	0.0745947479	
## LDA_01	0.0549743888	0.0080584702	
## LDA_02	-0.0814757616	0.0162884613	
## LDA_03	0.0037946292	-0.1828162993	
## LDA_04	-0.0002650426	0.0911846573	
## global_subjectivity	0.4646308125	0.6518911885	
## global_sentiment_polarity	0.1708459803	0.2088270873	
## global_rate_positive_words	0.2814757038	0.3850456166	
## global_rate_negative_words	0.1595003710	0.2599830942	
## rate_positive_words	0.4337401639	0.6132261845	
## rate_negative_words	0.1769621561	0.3114487045	

## avg_positive_polarity	0.4091535850	0.5803162333
## min_positive_polarity	0.4067620461	0.2273009637
## max_positive_polarity	0.1385059089	0.5187740665
## avg_negative_polarity	-0.2043636164	-0.3443621148
## min_negative_polarity	0.0438863610	-0.3025074331
## max_negative_polarity	-0.3281483679	-0.1916611838
## title_subjectivity	-0.0060146665	-0.0297809879
## title_sentiment_polarity	-0.0222721666	-0.0051729722
## abs_title_subjectivity	-0.0071953430	0.0178826252
## abs_title_sentiment_polarity	-0.0180250710	-0.0269930140
## shares	-0.0490997143	-0.0131875543
##	n_non_stop_unique_tokens	num_hrefs
## timedelta	0.178275135	4.534453e-04
## n_tokens_title	-0.048124646	-4.207178e-02
## n_tokens_content	-0.238593623	4.269887e-01
## n_unique_tokens	0.937127572	-1.214052e-01
## n_non_stop_words	0.757131834	1.656509e-01
## n_non_stop_unique_tokens	1.000000000	-1.128437e-01
## num_hrefs	-0.112843729	1.000000e+00
## num_self_hrefs	-0.031421499	3.976683e-01
## num_imgs	-0.298722288	3.432289e-01
## num_videos	0.009469396	1.090661e-01
## average_token_length	0.697327843	2.183807e-01
## num_keywords	-0.075369490	1.258921e-01
## data_channel_is_lifestyle	0.016686364	5.564531e-02
## data_channel_is_entertainment	-0.010479423	-3.568824e-03
## data_channel_is_bus	0.085130751	-6.376783e-02
## data_channel_is_socmed	0.017066328	6.431919e-02
## data_channel_is_tech	0.028773044	-6.162654e-02
## data_channel_is_world	-0.027647938	-2.435583e-02
## kw_min_min	0.082067148	-3.759106e-02
## kw_max_min	-0.001462669	1.444033e-02
## kw_avg_min	0.016341963	9.926370e-03
## kw_min_max	-0.017381905	-2.223239e-02
## kw_max_max	-0.087436900	4.805716e-02
## kw_avg_max	-0.092704114	-2.824547e-02
## kw_min_avg	-0.064108639	5.352410e-02
## kw_max_avg	-0.035249919	6.158474e-02
## kw_avg_avg	-0.102606237	1.098660e-01
## self_reference_min_shares	0.034658629	-6.893628e-03
## self_reference_max_shares	0.022200455	7.626543e-02
## self_reference_avg_share	0.036525586	2.290850e-02
## weekday_is_monday	0.006804124	-7.728561e-03
## weekday_is_tuesday	0.002202863	-1.108360e-02
## weekday_is_wednesday	0.008573763	-3.182490e-02
## weekday_is_thursday	0.002741152	-1.330572e-02
## weekday_is_friday	0.013127867	-1.154536e-03
## weekday_is_saturday	-0.033875319	6.232632e-02
## weekday_is_sunday	-0.017571959	4.180671e-02
## is_weekend	-0.037383000	7.587953e-02
## LDA_00	0.083393919	-2.229645e-02
## LDA_01	0.042871757	-5.276468e-02
## LDA_02	-0.029490458	5.170904e-05
## LDA_03	-0.120612333	1.154260e-01
## LDA_04	0.040476368	-5.416315e-02
## global_subjectivity	0.505801442	1.986004e-01

## global_sentiment_polarity	0.170137588	8.860376e-02
## global_rate_positive_words	0.333296797	6.029355e-02
## global_rate_negative_words	0.206277268	2.964742e-02
## rate_positive_words	0.478025612	1.011172e-01
## rate_negative_words	0.218535374	5.265140e-02
## avg_positive_polarity	0.448698492	1.832535e-01
## min_positive_polarity	0.320486805	-9.036846e-02
## max_positive_polarity	0.257142494	2.862188e-01
## avg_negative_polarity	-0.243354934	-1.487751e-01
## min_negative_polarity	-0.082220939	-2.609270e-01
## max_negative_polarity	-0.259059387	5.841709e-02
## title_subjectivity	-0.027689561	3.790168e-02
## title_sentiment_polarity	-0.030829201	3.526619e-02
## abs_title_subjectivity	0.009043539	9.212141e-03
## abs_title_sentiment_polarity	-0.036858646	5.026503e-02
## shares	-0.050807230	7.765240e-02
##	num_self_hrefs	num_imgs
## timedelta	0.0682016912	-0.0205447979
## n_tokens_title	-0.0065999986	0.0005176119
## n_tokens_content	0.3036560429	0.3570744479
## n_unique_tokens	-0.0596678473	-0.2466785971
## n_non_stop_words	0.1444486919	0.0121732380
## n_non_stop_unique_tokens	-0.0314214988	-0.2987222876
## num_hrefs	0.3976682755	0.3432288694
## num_self_hrefs	1.0000000000	0.2520156229
## num_imgs	0.2520156229	1.0000000000
## num_videos	0.0705273875	-0.0612288735
## average_token_length	0.1183863198	0.0272873219
## num_keywords	0.1015440757	0.0809250254
## data_channel_is_lifestyle	-0.0461519349	0.0081955585
## data_channel_is_entertainment	0.0257414954	0.1081518444
## data_channel_is_bus	-0.0605801933	-0.1409800310
## data_channel_is_socmed	0.0963244572	0.0014164457
## data_channel_is_tech	0.1722219573	0.0015597361
## data_channel_is_world	-0.1237703238	-0.1074752672
## kw_min_min	-0.0098363776	-0.0365145710
## kw_max_min	0.0010843173	0.0019098312
## kw_avg_min	0.0002531275	-0.0045073296
## kw_min_max	-0.0269503034	-0.0023494404
## kw_max_max	0.0020657210	0.0440200488
## kw_avg_max	-0.0656222233	-0.0048209735
## kw_min_avg	0.0396793993	0.0756982425
## kw_max_avg	-0.0189238183	0.0528947757
## kw_avg_avg	-0.0178151848	0.1302361730
## self_reference_min_shares	-0.0291000819	0.0024468750
## self_reference_max_shares	0.1228662357	0.0359680200
## self_reference_avg_shares	0.0219408481	0.0179421804
## weekday_is_monday	0.0120801622	-0.0052795232
## weekday_is_tuesday	0.0015388713	-0.0051657269
## weekday_is_wednesday	-0.0189077636	-0.0220196592
## weekday_is_thursday	-0.0172978510	-0.0076521440
## weekday_is_friday	-0.0224442586	-0.0080357424
## weekday_is_saturday	0.0506180573	0.0351169993
## weekday_is_sunday	0.0195431350	0.0403366565
## is_weekend	0.0508297019	0.0553253322
## LDA_00	-0.0272265624	-0.1477662824

## LDA_01	-0.0016481582	0.0572390885	0.0188465189
## LDA_02	-0.0889478888	-0.0887089049	-0.0907721417
## LDA_03	-0.0250618535	0.1970160873	0.2724511483
## LDA_04	0.1396231473	-0.0161391512	-0.1270314986
## global_subjectivity	0.1137495748	0.0764689844	0.0804586919
## global_sentiment_polarity	0.0919807964	0.0235957300	-0.0324520654
## global_rate_positive_words	0.1217489782	-0.0332051985	0.0734856197
## global_rate_negative_words	0.0095762196	0.0219303838	0.1857185419
## rate_positive_words	0.1394031604	-0.0176314546	-0.0469267424
## rate_negative_words	-0.0160004545	0.0343617682	0.0738696730
## avg_positive_polarity	0.0949576139	0.0890817780	0.0977056701
## min_positive_polarity	-0.0736776606	-0.0344629463	-0.0139299326
## max_positive_polarity	0.1868157224	0.1582220732	0.1247848854
## avg_negative_polarity	-0.0552010441	-0.0675117578	-0.1189243120
## min_negative_polarity	-0.1175609041	-0.1353264489	-0.1417098512
## max_negative_polarity	0.0403527772	0.0473107796	0.0297290417
## title_subjectivity	-0.0115076057	0.0443396311	0.0632949524
## title_sentiment_polarity	0.0248993025	0.0447434528	0.0163833176
## abs_title_subjectivity	0.0086652061	-0.0096679951	-0.0263412425
## abs_title_sentiment_polarity	-0.0104585510	0.0544530836	0.0539458804
## shares	0.0431620836	0.0559268338	-0.0028983733
##	average_token_length	num_keywords	
## timedelta	1.192985e-01	0.0559806387	
## n_tokens_title	-6.779093e-02	-0.0015221627	
## n_tokens_content	1.642624e-01	0.0706915910	
## n_unique_tokens	6.435803e-01	-0.0749778439	
## n_non_stop_words	9.407069e-01	-0.0235368421	
## n_non_stop_unique_tokens	6.973278e-01	-0.0753694899	
## num_hrefs	2.183807e-01	0.1258921289	
## num_self_hrefs	1.183863e-01	0.1015440757	
## num_imgs	2.728732e-02	0.0809250254	
## num_videos	-8.241554e-03	-0.0259140081	
## average_token_length	1.000000e+00	-0.0156899937	
## num_keywords	-1.568999e-02	1.0000000000	
## data_channel_is_lifestyle	1.093580e-02	0.1236777352	
## data_channel_is_entertainment	-4.747738e-02	-0.0750124191	
## data_channel_is_bus	7.138103e-02	-0.1734609250	
## data_channel_is_socmed	2.268430e-02	-0.0872695696	
## data_channel_is_tech	1.320286e-02	0.1438745201	
## data_channel_is_world	9.127962e-02	0.0202426576	
## kw_min_min	3.097451e-02	0.0036223713	
## kw_max_min	1.613608e-03	0.0877423985	
## kw_avg_min	1.492220e-02	0.0884672002	
## kw_min_max	-4.158466e-02	-0.2810847335	
## kw_max_max	-3.481498e-02	0.0044623144	
## kw_avg_max	-1.510131e-01	-0.3407158615	
## kw_min_avg	-7.542085e-02	-0.2821125653	
## kw_max_avg	-4.572101e-02	0.1224896207	
## kw_avg_avg	-1.424489e-01	0.0211787854	
## self_reference_min_shares	2.767052e-02	-0.0024914632	
## self_reference_max_shares	3.405127e-02	0.0140837080	
## self_reference_avg_shares	3.551293e-02	0.0062724930	
## weekday_is_monday	1.104076e-03	-0.0214825988	
## weekday_is_tuesday	-2.145008e-03	-0.0093379414	
## weekday_is_wednesday	-3.834284e-03	-0.0189137480	
## weekday_is_thursday	3.834747e-03	-0.0116766354	



```

## weekday_is_friday      8.560885e-05  0.0014335234
## weekday_is_saturday    -5.596734e-03  0.0417842304
## weekday_is_sunday      6.972856e-03  0.0527869859
## is_weekend             1.222553e-03  0.0694193691
## LDA_00                 7.954466e-02 -0.1587356221
## LDA_01                 -2.851828e-02 -0.0629639382
## LDA_02                 1.071300e-01 -0.0152822508
## LDA_03                 -1.877577e-01  0.0193298459
## LDA_04                 2.891304e-02  0.1887555507
## global_subjectivity    5.764047e-01  0.0416078185
## global_sentiment_polarity 1.674536e-01  0.0870387629
## global_rate_positive_words 3.060818e-01  0.0533543243
## global_rate_negative_words 2.182558e-01 -0.0423055704
## rate_positive_words    5.611272e-01  0.0427800847
## rate_negative_words    3.121484e-01 -0.0754916333
## avg_positive_polarity  5.215016e-01  0.0337669405
## min_positive_polarity  2.114862e-01 -0.0161886066
## max_positive_polarity  4.513832e-01  0.0630475224
## avg_negative_polarity -3.105721e-01  0.0234604734
## min_negative_polarity -2.660084e-01  0.0006686681
## max_negative_polarity -1.869905e-01  0.0270013290
## title_subjectivity     -4.105006e-02  0.0142986506
## title_sentiment_polarity -1.295987e-02  0.0292319340
## abs_title_subjectivity  2.671961e-02 -0.0095351693
## abs_title_sentiment_polarity -3.640478e-02  0.0195806869
## shares                 -2.555183e-02  0.0655351664
## data_channel_is_lifestyle
## timedelta              0.056793320
## n_tokens_title          -0.070580414
## n_tokens_content        0.032286035
## n_unique_tokens         -0.010447984
## n_non_stop_words        0.028131702
## n_non_stop_unique_tokens 0.016686364
## num_hrefs               0.055645308
## num_self_hrefs          -0.046151935
## num_imgs                0.008195559
## num_videos              -0.043296116
## average_token_length    0.010935798
## num_keywords            0.123677735
## data_channel_is_lifestyle 1.000000000
## data_channel_is_entertainment -0.109159148
## data_channel_is_bus      -0.102825927
## data_channel_is_socmed   -0.056944488
## data_channel_is_tech     -0.111274113
## data_channel_is_world    -0.124986806
## kw_min_min              0.052380920
## kw_max_min              0.031779929
## kw_avg_min              0.042750478
## kw_min_max              -0.025971322
## kw_max_max              -0.055843827
## kw_avg_max              -0.131250678
## kw_min_avg              -0.010211378
## kw_max_avg              0.044138036
## kw_avg_avg              0.057864352
## self_reference_min_shares 0.007174742
## self_reference_max_shares -0.010051891

```

## self_reference_avg_sharess	-0.001368446	
## weekday_is_monday	-0.008784842	
## weekday_is_tuesday	-0.015840227	
## weekday_is_wednesday	-0.001304170	
## weekday_is_thursday	-0.008977192	
## weekday_is_friday	0.002573411	
## weekday_is_saturday	0.023781378	
## weekday_is_sunday	0.027660613	
## is_weekend	0.037724605	
## LDA_00	-0.008564076	
## LDA_01	-0.079047669	
## LDA_02	-0.118612958	
## LDA_03	-0.058244926	
## LDA_04	0.244100287	
## global_subjectivity	0.064573555	
## global_sentiment_polarity	0.082799358	
## global_rate_positive_words	0.066145384	
## global_rate_negative_words	-0.008731043	
## rate_positive_words	0.053844760	
## rate_negative_words	-0.035134063	
## avg_positive_polarity	0.068624529	
## min_positive_polarity	-0.003067211	
## max_positive_polarity	0.068905301	
## avg_negative_polarity	-0.007507156	
## min_negative_polarity	-0.022334144	
## max_negative_polarity	0.008891760	
## title_subjectivity	0.004311771	
## title_sentiment_polarity	0.039277937	
## abs_title_subjectivity	0.005768648	
## abs_title_sentiment_polarity	0.021135729	
## shares	0.031436922	
##	data_channel_is_entertainment	data_channel_is_bus
## timedelta	-0.0495322721	0.0625109118
## n_tokens_title	0.1346360586	-0.0285552881
## n_tokens_content	0.0649442760	-0.0182403977
## n_unique_tokens	0.0012134419	0.0518389810
## n_non_stop_words	-0.0007286326	0.0647419326
## n_non_stop_unique_tokens	-0.0104794230	0.0851307510
## num_hrefs	-0.0035688238	-0.0637678283
## num_self_hrefs	0.0257414954	-0.0605801933
## num_imgs	0.1081518444	-0.1409800310
## num_videos	0.1588252043	-0.0710138102
## average_token_length	-0.0474773842	0.0713810294
## num_keywords	-0.0750124191	-0.1734609250
## data_channel_is_lifestyle	-0.1091591478	-0.1028259271
## data_channel_is_entertainment	1.0000000000	-0.2079134194
## data_channel_is_bus	-0.2079134194	1.0000000000
## data_channel_is_socmed	-0.1151414200	-0.1084611185
## data_channel_is_tech	-0.2249956022	-0.2119417554
## data_channel_is_world	-0.2527225853	-0.2380600680
## kw_min_min	-0.0331596904	0.0235509968
## kw_max_min	-0.0127507194	-0.0190511364
## kw_avg_min	-0.0213645326	-0.0007731041
## kw_min_max	0.0002008767	0.0496055175
## kw_max_max	0.0312211027	-0.0219969043
## kw_avg_max	-0.0485045516	0.1887624947

## kw_min_avg	0.0125557578	-0.0006964724
## kw_max_avg	0.0026382069	-0.0332249145
## kw_avg_avg	0.0192219465	-0.0606934748
## self_reference_min_shares	-0.0281141539	-0.0108034163
## self_reference_max_shares	-0.0143367963	-0.0036328929
## self_reference_avg_sharess	-0.0241424920	-0.0080424789
## weekday_is_monday	0.0312982483	0.0182885616
## weekday_is_tuesday	-0.0032272333	0.0011865906
## weekday_is_wednesday	-0.0026542661	0.0175014155
## weekday_is_thursday	-0.0118999459	0.0178937872
## weekday_is_friday	-0.0090069153	-0.0127654453
## weekday_is_saturday	-0.0149546591	-0.0450055650
## weekday_is_sunday	0.0079076618	-0.0236520231
## is_weekend	-0.0047674210	-0.0498953915
## LDA_00	-0.2125885414	0.7844410589
## LDA_01	0.6029099749	-0.1293727477
## LDA_02	-0.2178212353	-0.2216885820
## LDA_03	0.2319348114	-0.2255954454
## LDA_04	-0.2787902043	-0.1717509692
## global_subjectivity	0.0387047894	-0.0283848381
## global_sentiment_polarity	-0.0375126143	0.0758566798
## global_rate_positive_words	0.0290134218	0.0870344016
## global_rate_negative_words	0.1108850476	-0.0780983479
## rate_positive_words	-0.0419899240	0.1264803995
## rate_negative_words	0.0501451079	-0.0849968772
## avg_positive_polarity	0.0556222560	-0.0003380262
## min_positive_polarity	-0.0163018590	-0.0493607733
## max_positive_polarity	0.0811859377	0.0195035875
## avg_negative_polarity	-0.1340331838	0.0608214315
## min_negative_polarity	-0.1115385523	0.0704959576
## max_negative_polarity	-0.0140874536	-0.0097482046
## title_subjectivity	0.0542389183	-0.0421391066
## title_sentiment_polarity	-0.0093056757	0.0190566846
## abs_title_subjectivity	-0.0498677520	-0.0049787787
## abs_title_sentiment_polarity	0.0386364693	-0.0291278490
## shares	-0.1054217522	0.0016397432
##	data_channel_is_socmed	data_channel_is_tech
## timedelta	0.073001375	0.083176559
## n_tokens_title	-0.087747916	-0.043980691
## n_tokens_content	0.028435253	0.017566382
## n_unique_tokens	0.013369665	0.004464515
## n_non_stop_words	0.033046468	0.073595395
## n_non_stop_unique_tokens	0.017066328	0.028773044
## num_hrefs	0.064319192	-0.061626539
## num_self_hrefs	0.096324457	0.172221957
## num_imgs	0.001416446	0.001559736
## num_videos	-0.006762154	-0.092544899
## average_token_length	0.022684300	0.013202855
## num_keywords	-0.087269570	0.143874520
## data_channel_is_lifestyle	-0.056944488	-0.111274113
## data_channel_is_entertainment	-0.115141420	-0.224995602
## data_channel_is_bus	-0.108461118	-0.211941755
## data_channel_is_socmed	1.000000000	-0.117372292
## data_channel_is_tech	-0.117372292	1.000000000
## data_channel_is_world	-0.131836484	-0.257619102
## kw_min_min	0.034689348	0.021743658

## kw_max_min	0.001211095	0.004206577	
## kw_avg_min	0.027572210	0.008825899	
## kw_min_max	0.065015450	-0.051834800	
## kw_max_max	-0.035897335	-0.030519461	
## kw_avg_max	-0.053610193	-0.152108045	
## kw_min_avg	0.044635330	-0.048410335	
## kw_max_avg	-0.011974928	-0.079162943	
## kw_avg_avg	0.020864466	-0.130008834	
## self_reference_min_shares	0.010256451	0.014619017	
## self_reference_max_shares	0.028837600	0.024504363	
## self_reference_avg_shareess	0.018205539	0.022685427	
## weekday_is_monday	-0.016603816	0.002544870	
## weekday_is_tuesday	0.008018744	0.015770755	
## weekday_is_wednesday	-0.005055217	0.002757717	
## weekday_is_thursday	0.010036668	-0.003927026	
## weekday_is_friday	-0.004069481	-0.009565629	
## weekday_is_saturday	0.021900246	0.018074582	
## weekday_is_sunday	-0.010561086	-0.030460328	
## is_weekend	0.007745195	-0.009897804	
## LDA_00	0.175165930	-0.203875788	
## LDA_01	-0.065854772	-0.165134337	
## LDA_02	-0.020802377	-0.192860984	
## LDA_03	-0.022639058	-0.248447906	
## LDA_04	-0.066339915	0.750710266	
## global_subjectivity	0.040181736	0.062245880	
## global_sentiment_polarity	0.073124886	0.136115521	
## global_rate_positive_words	0.098437937	0.092712855	
## global_rate_negative_words	-0.020997239	-0.097941625	
## rate_positive_words	0.080617583	0.163248806	
## rate_negative_words	-0.062995045	-0.118966547	
## avg_positive_polarity	0.017530561	0.019072389	
## min_positive_polarity	-0.049185190	0.030301806	
## max_positive_polarity	0.028126015	0.028763303	
## avg_negative_polarity	0.006036885	0.114139446	
## min_negative_polarity	0.007020899	0.119111689	
## max_negative_polarity	-0.013101920	0.017147486	
## title_subjectivity	-0.014969652	-0.041053663	
## title_sentiment_polarity	0.026065901	0.026762990	
## abs_title_subjectivity	0.011879254	0.007123074	
## abs_title_sentiment_polarity	-0.002916347	-0.042364623	
## shares	0.114944422	0.097379147	
##	data_channel_is_world	kw_min_min	kw_max_min
## timedelta	-0.1770178793	0.591729555	0.0324452683
## n_tokens_title	0.0464948750	-0.106240706	-0.0047284594
## n_tokens_content	0.0601794864	-0.057375203	0.0027778389
## n_unique_tokens	-0.0773288219	0.091382122	-0.0017722446
## n_non_stop_words	-0.0005446489	0.048114619	0.0004074634
## n_non_stop_unique_tokens	-0.0276479384	0.082067148	-0.0014626690
## num_hrefs	-0.0243558280	-0.037591059	0.0144403282
## num_self_hrefs	-0.1237703238	-0.009836378	0.0010843173
## num_imgs	-0.1074752672	-0.036514571	0.0019098312
## num_videos	-0.0890937022	0.016429591	0.0028181494
## average_token_length	0.0912796152	0.030974510	0.0016136075
## num_keywords	0.0202426576	0.003622371	0.0877423985
## data_channel_is_lifestyle	-0.1249868057	0.052380920	0.0317799289
## data_channel_is_entertainment	-0.2527225853	-0.033159690	-0.0127507194

## data_channel_is_bus	-0.2380600680	0.023550997	-0.0190511364
## data_channel_is_socmed	-0.1318364844	0.034689348	0.0012110952
## data_channel_is_tech	-0.2576191018	0.021743658	0.0042065767
## data_channel_is_world	1.0000000000	-0.081809828	-0.0155132510
## kw_min_min	-0.0818098281	1.0000000000	0.0145213745
## kw_max_min	-0.0155132510	0.014521375	1.0000000000
## kw_avg_min	-0.0390513865	0.127537234	0.9346521551
## kw_min_max	-0.0603840070	-0.061235993	-0.0402816792
## kw_max_max	0.0861884419	-0.859348713	-0.0078961426
## kw_avg_max	-0.1141767361	-0.505503655	-0.0441113463
## kw_min_avg	-0.1250152427	-0.134986115	0.0028764153
## kw_max_avg	-0.0758382841	-0.071101754	0.5755953307
## kw_avg_avg	-0.2436930358	-0.182431258	0.3888656963
## self_reference_min_shares	-0.0246965036	-0.021281456	0.0124878242
## self_reference_max_shares	-0.0498940715	-0.038822604	0.0502231070
## self_reference_avg_shareess	-0.0438529083	-0.035812209	0.0364095667
## weekday_is_monday	-0.0117207610	0.004022713	0.0076006456
## weekday_is_tuesday	-0.0048686518	-0.006769669	-0.0028911764
## weekday_is_wednesday	-0.0014840778	0.004010434	0.0007276445
## weekday_is_thursday	0.0014082204	0.004663634	0.0005834359
## weekday_is_friday	0.0184087123	0.001917593	0.0002616779
## weekday_is_saturday	0.0002266310	-0.020078565	-0.0042214896
## weekday_is_sunday	-0.0008924224	0.007463018	-0.0052426390
## is_weekend	-0.0005065492	-0.008763790	-0.0069457120
## LDA_00	-0.2453241492	0.029936296	-0.0055070296
## LDA_01	-0.2139233094	0.007278328	-0.0031861212
## LDA_02	0.8412514069	-0.079613810	-0.0144927663
## LDA_03	-0.2648245985	0.003268935	0.0190518628
## LDA_04	-0.1886681719	0.043253280	0.0029700462
## global_subjectivity	-0.1808165308	0.040822516	0.0139624549
## global_sentiment_polarity	-0.2420297143	0.077736777	0.0109083135
## global_rate_positive_words	-0.2579145127	0.097657471	0.0100678123
## global_rate_negative_words	0.0258024368	-0.003240620	-0.0013263840
## rate_positive_words	-0.1716632396	0.081587396	0.0074614217
## rate_negative_words	0.2065499263	-0.047471333	-0.0084590926
## avg_positive_polarity	-0.1477883422	0.044526497	0.0107667066
## min_positive_polarity	-0.0498410041	0.023058385	0.0044944052
## max_positive_polarity	-0.1239435828	0.021482027	0.0130237075
## avg_negative_polarity	0.0218378591	0.016757468	-0.0065193842
## min_negative_polarity	-0.0802036440	0.051137660	-0.0032133776
## max_negative_polarity	0.0754476879	-0.027942583	-0.0016062640
## title_subjectivity	-0.0606835036	-0.003169693	0.0118393231
## title_sentiment_polarity	-0.0853572063	0.032088210	0.0152019338
## abs_title_subjectivity	0.0564847711	0.006509745	-0.0030479510
## abs_title_sentiment_polarity	-0.0643434945	0.009404651	0.0158622739
## shares	-0.1374309832	0.039892832	0.0224717500
##	kw_avg_min	kw_min_max	kw_max_max
## timedelta	0.1475609966	-0.0770000797	-0.6399255709
## n_tokens_title	-0.0340093710	0.0105796801	0.1184967367
## n_tokens_content	-0.0019092958	-0.0223714013	0.0636262772
## n_unique_tokens	0.0139020688	-0.0052617458	-0.0978279113
## n_non_stop_words	0.0179855305	-0.0403968837	-0.0520226276
## n_non_stop_unique_tokens	0.0163419634	-0.0173819054	-0.0874369001
## num_hrefs	0.0099263697	-0.0222323850	0.0480571627
## num_self_hrefs	0.0002531275	-0.0269503034	0.0020657210
## num_imgs	-0.0045073296	-0.0023494404	0.0440200488

## num_videos	-0.0069325466	0.0429232512	-0.0042290740
## average_token_length	0.0149221959	-0.0415846602	-0.0348149842
## num_keywords	0.0884672002	-0.2810847335	0.0044623144
## data_channel_is_lifestyle	0.0427504781	-0.0259713219	-0.0558438266
## data_channel_is_entertainment	-0.0213645326	0.0002008767	0.0312211027
## data_channel_is_bus	-0.0007731041	0.0496055175	-0.0219969043
## data_channel_is_socmed	0.0275722100	0.0650154498	-0.0358973352
## data_channel_is_tech	0.0088258988	-0.0518348002	-0.0305194614
## data_channel_is_world	-0.0390513865	-0.0603840070	0.0861884419
## kw_min_min	0.1275372341	-0.0612359925	-0.8593487127
## kw_max_min	0.9346521551	-0.0402816792	-0.0078961426
## kw_avg_min	1.0000000000	-0.0703096193	-0.1208960043
## kw_min_max	-0.0703096193	1.0000000000	0.0655826373
## kw_max_max	-0.1208960043	0.0655826373	1.0000000000
## kw_avg_max	-0.1488097059	0.4154691778	0.5662156916
## kw_min_avg	-0.0295082852	0.3548511320	0.1565144812
## kw_max_avg	0.5260782595	0.0235316459	0.0925306112
## kw_avg_avg	0.3609997150	0.1730671767	0.2282184774
## self_reference_min_shares	0.0126927612	0.0164271356	0.0280293227
## self_reference_max_shares	0.0389491094	0.0210560718	0.0428894164
## self_reference_avg_shares	0.0310050669	0.0204389785	0.0423276552
## weekday_is_monday	0.0029297585	-0.0133045154	-0.0103852891
## weekday_is_tuesday	-0.0035470265	0.0004519030	0.0059012443
## weekday_is_wednesday	0.0007906250	0.0090946148	-0.0099909585
## weekday_is_thursday	0.0008680735	0.0036309300	-0.0051739833
## weekday_is_friday	0.0036080029	-0.0032346500	0.0035699056
## weekday_is_saturday	-0.0049659708	0.0076883591	0.0234798507
## weekday_is_sunday	-0.0017793343	-0.0035150584	0.0026210114
## is_weekend	-0.0048833655	0.0028632976	0.0187501317
## LDA_00	0.0212850965	0.0526944741	-0.0293138889
## LDA_01	0.0003072777	-0.0106903828	-0.0024778594
## LDA_02	-0.0333362324	-0.0562326102	0.0781220615
## LDA_03	0.0016371483	0.0736896811	0.0002643329
## LDA_04	0.0119533857	-0.0570562071	-0.0495010940
## global_subjectivity	0.0277590138	-0.0176456187	-0.0440194311
## global_sentiment_polarity	0.0317358159	-0.0126078166	-0.0877603215
## global_rate_positive_words	0.0368277185	-0.0104696172	-0.1079100698
## global_rate_negative_words	-0.0025277088	0.0013907735	0.0077529910
## rate_positive_words	0.0317094945	-0.0267523586	-0.0893983023
## rate_negative_words	-0.0192019399	-0.0103569641	0.0529117278
## avg_positive_polarity	0.0234948729	-0.0099704448	-0.0478631076
## min_positive_polarity	0.0053577922	0.0012941374	-0.0249162804
## max_positive_polarity	0.0250246573	-0.0205172098	-0.0212068689
## avg_negative_polarity	-0.0042679315	-0.0046766467	-0.0181866391
## min_negative_polarity	0.0032159007	0.0018909466	-0.0560416289
## max_negative_polarity	-0.0072695900	-0.0066928923	0.0295382422
## title_subjectivity	0.0072770474	0.0205257434	0.0041814762
## title_sentiment_polarity	0.0171955749	0.0090981265	-0.0296184875
## abs_title_subjectivity	0.0019349306	-0.0100669266	-0.0080034279
## abs_title_sentiment_polarity	0.0118642423	0.0219397266	-0.0069176430
## shares	0.0316292070	0.0078409494	-0.0249163874
##	kw_avg_max	kw_min_avg	kw_max_avg
## timedelta	-0.494588393	-0.1572421960	-0.0518910395
## n_tokens_title	0.105279531	-0.0020498158	0.0062752369
## n_tokens_content	-0.090259099	-0.0271703033	-0.0278774252
## n_unique_tokens	-0.045292568	-0.0253407244	-0.0073578140

## n_non_stop_words	-0.165601972	-0.0741840808	-0.0470906798
## n_non_stop_unique_tokens	-0.092704114	-0.0641086395	-0.0352499191
## num_hrefs	-0.028245470	0.0535241043	0.0615847435
## num_self_hrefs	-0.065622223	0.0396793993	-0.0189238183
## num_imgs	-0.004820973	0.0756982425	0.0528947757
## num_videos	0.128167721	0.0584250428	0.0512694388
## average_token_length	-0.151013099	-0.0754208485	-0.0457210122
## num_keywords	-0.340715861	-0.2821125653	0.1224896207
## data_channel_is_lifestyle	-0.131250678	-0.0102113782	0.0441380356
## data_channel_is_entertainment	-0.048504552	0.0125557578	0.0026382069
## data_channel_is_bus	0.188762495	-0.0006964724	-0.0332249145
## data_channel_is_socmed	-0.053610193	0.0446353302	-0.0119749283
## data_channel_is_tech	-0.152108045	-0.0484103346	-0.0791629432
## data_channel_is_world	-0.114176736	-0.1250152427	-0.0758382841
## kw_min_min	-0.505503655	-0.1349861148	-0.0711017538
## kw_max_min	-0.044111346	0.0028764153	0.5755953307
## kw_avg_min	-0.148809706	-0.0295082852	0.5260782595
## kw_min_max	0.415469178	0.3548511320	0.0235316459
## kw_max_max	0.566215692	0.1565144812	0.0925306112
## kw_avg_max	1.000000000	0.4122648459	0.1358134849
## kw_min_avg	0.412264846	1.0000000000	0.0914122908
## kw_max_avg	0.135813485	0.0914122908	1.0000000000
## kw_avg_avg	0.431416915	0.4624615943	0.7974214999
## self_reference_min_shares	0.050041121	0.0320728094	0.0716087125
## self_reference_max_shares	0.085744222	0.0530782601	0.1706100701
## self_reference_avg_share	0.082253311	0.0508899107	0.1501447521
## weekday_is_monday	-0.002909271	-0.0147549811	-0.0083437160
## weekday_is_tuesday	0.006885927	-0.0045049883	-0.0068649856
## weekday_is_wednesday	0.004820018	-0.0127289654	-0.0044782944
## weekday_is_thursday	0.009335798	-0.0092367649	0.0004508515
## weekday_is_friday	0.003380585	-0.0014379645	0.0006307004
## weekday_is_saturday	-0.006387125	0.0390371529	0.0164976482
## weekday_is_sunday	-0.027224193	0.0283060814	0.0129341380
## is_weekend	-0.024961905	0.0491150467	0.0214845937
## LDA_00	0.128835155	0.0041880270	-0.0174753201
## LDA_01	-0.076631871	-0.0087923884	-0.0006440454
## LDA_02	-0.112363713	-0.1166254432	-0.0992203791
## LDA_03	0.274437862	0.1744846909	0.1980507015
## LDA_04	-0.219697321	-0.0541945968	-0.0814010865
## global_subjectivity	-0.065534423	0.0089891255	0.0408839863
## global_sentiment_polarity	-0.083562480	-0.0010813659	0.0132591478
## global_rate_positive_words	-0.105812493	-0.0052990783	0.0010372841
## global_rate_negative_words	0.003378989	0.0169412988	0.0132150395
## rate_positive_words	-0.145164002	-0.0564006042	-0.0339006680
## rate_negative_words	0.000302288	-0.0107983963	-0.0085407047
## avg_positive_polarity	-0.069275315	0.0085896859	0.0298279689
## min_positive_polarity	0.006240925	0.0199308803	0.0197383489
## max_positive_polarity	-0.097713160	-0.0152242894	0.0093002865
## avg_negative_polarity	-0.018997532	-0.0192332438	-0.0454553715
## min_negative_polarity	0.013644325	-0.0006702353	-0.0230615865
## max_negative_polarity	-0.021596146	-0.0112862262	-0.0221019680
## title_subjectivity	0.044194274	0.0419655849	0.0433899106
## title_sentiment_polarity	-0.011962494	0.0264624061	0.0210335125
## abs_title_subjectivity	-0.013602757	-0.0151542088	-0.0130944317
## abs_title_sentiment_polarity	0.032303890	0.0536614354	0.0469152925
## shares	0.016024753	0.0895102100	0.0631574499

##	kw_avg_avg	self_reference_min_shares
## timedelta	-0.1640704979	-1.332549e-02
## n_tokens_title	0.0028811365	9.422205e-05
## n_tokens_content	-0.0789510010	-2.672358e-02
## n_unique_tokens	-0.0299799112	4.105054e-02
## n_non_stop_words	-0.1422126841	3.192904e-02
## n_non_stop_unique_tokens	-0.1026062374	3.465863e-02
## num_hrefs	0.1098659925	-6.893628e-03
## num_self_hrefs	-0.0178151848	-2.910008e-02
## num_imgs	0.1302361730	2.446875e-03
## num_videos	0.1127285096	-1.556014e-03
## average_token_length	-0.1424488609	2.767052e-02
## num_keywords	0.0211787854	-2.491463e-03
## data_channel_is_lifestyle	0.0578643524	7.174742e-03
## data_channel_is_entertainment	0.0192219465	-2.811415e-02
## data_channel_is_bus	-0.0606934748	-1.080342e-02
## data_channel_is_socmed	0.0208644664	1.025645e-02
## data_channel_is_tech	-0.1300088340	1.461902e-02
## data_channel_is_world	-0.2436930358	-2.469650e-02
## kw_min_min	-0.1824312575	-2.128146e-02
## kw_max_min	0.3888656963	1.248782e-02
## kw_avg_min	0.3609997150	1.269276e-02
## kw_min_max	0.1730671767	1.642714e-02
## kw_max_max	0.2282184774	2.802932e-02
## kw_avg_max	0.4314169154	5.004112e-02
## kw_min_avg	0.4624615943	3.207281e-02
## kw_max_avg	0.7974214999	7.160871e-02
## kw_avg_avg	1.0000000000	8.542298e-02
## self_reference_min_shares	0.0854229820	1.000000e+00
## self_reference_max_shares	0.1663929507	4.848502e-01
## self_reference_avg_share	0.1534354634	8.198209e-01
## weekday_is_monday	-0.0234931415	-4.730205e-03
## weekday_is_tuesday	-0.0060728009	-3.192909e-03
## weekday_is_wednesday	-0.0133990764	-1.357637e-03
## weekday_is_thursday	-0.0041020941	4.024969e-03
## weekday_is_friday	0.0046796872	1.691445e-03
## weekday_is_saturday	0.0369802184	-2.821745e-03
## weekday_is_sunday	0.0302187109	8.243641e-03
## is_weekend	0.0490773571	4.158521e-03
## LDA_00	-0.0337448455	-2.396377e-03
## LDA_01	0.0154544313	-1.347316e-02
## LDA_02	-0.2709730674	-3.274662e-02
## LDA_03	0.4328770857	3.017573e-02
## LDA_04	-0.1408983296	1.511471e-02
## global_subjectivity	0.0637427343	4.995306e-02
## global_sentiment_polarity	0.0222001707	6.052995e-03
## global_rate_positive_words	-0.0005482842	1.941381e-03
## global_rate_negative_words	0.0259119415	1.103102e-02
## rate_positive_words	-0.0931143175	1.880080e-02
## rate_negative_words	-0.0374691238	1.096129e-02
## avg_positive_polarity	0.0404843291	2.774315e-02
## min_positive_polarity	0.0484266071	2.551053e-02
## max_positive_polarity	-0.0068446760	2.522100e-03
## avg_negative_polarity	-0.0657919748	-3.062984e-02
## min_negative_polarity	-0.0132621296	-4.781067e-03
## max_negative_polarity	-0.0446349077	-4.312440e-02



## title_subjectivity	0.0885870699	2.078068e-03
## title_sentiment_polarity	0.0365002622	-7.524291e-04
## abs_title_subjectivity	-0.0242812137	3.229158e-03
## abs_title_sentiment_polarity	0.0933080525	4.142497e-03
## shares	0.1476775609	4.458771e-02
##	self_reference_max_shares	
## timedelta	-0.015608824	
## n_tokens_title	0.001031335	
## n_tokens_content	0.026634429	
## n_unique_tokens	0.022774861	
## n_non_stop_words	0.040886686	
## n_non_stop_unique_tokens	0.022200455	
## num_hrefs	0.076265430	
## num_self_hrefs	0.122866236	
## num_imgs	0.035968020	
## num_videos	0.068851960	
## average_token_length	0.034051269	
## num_keywords	0.014083708	
## data_channel_is_lifestyle	-0.010051891	
## data_channel_is_entertainment	-0.014336796	
## data_channel_is_bus	-0.003632893	
## data_channel_is_socmed	0.028837600	
## data_channel_is_tech	0.024504363	
## data_channel_is_world	-0.049894071	
## kw_min_min	-0.038822604	
## kw_max_min	0.050223107	
## kw_avg_min	0.038949109	
## kw_min_max	0.021056072	
## kw_max_max	0.042889416	
## kw_avg_max	0.085744222	
## kw_min_avg	0.053078260	
## kw_max_avg	0.170610070	
## kw_avg_avg	0.166392951	
## self_reference_min_shares	0.484850207	
## self_reference_max_shares	1.000000000	
## self_reference_avg_sharess	0.858243324	
## weekday_is_monday	-0.003787376	
## weekday_is_tuesday	-0.002456102	
## weekday_is_wednesday	0.003629530	
## weekday_is_thursday	-0.005178911	
## weekday_is_friday	0.003404696	
## weekday_is_saturday	0.002784834	
## weekday_is_sunday	0.004421065	
## is_weekend	0.005303100	
## LDA_00	0.008611267	
## LDA_01	-0.021078658	
## LDA_02	-0.042127866	
## LDA_03	0.041924681	
## LDA_04	0.008551555	
## global_subjectivity	0.060646061	
## global_sentiment_polarity	0.007417009	
## global_rate_positive_words	0.018306755	
## global_rate_negative_words	0.020009228	
## rate_positive_words	0.023059910	
## rate_negative_words	0.015031990	
## avg_positive_polarity	0.037903252	

## min_positive_polarity	-0.002398750	
## max_positive_polarity	0.037143277	
## avg_negative_polarity	-0.047057792	
## min_negative_polarity	-0.040224929	
## max_negative_polarity	-0.021758056	
## title_subjectivity	0.010066119	
## title_sentiment_polarity	0.002122743	
## abs_title_subjectivity	-0.003884220	
## abs_title_sentiment_polarity	0.007145344	
## shares	0.054805408	
##	self_reference_avg_shares	weekday_is_monday
## timedelta	-1.655431e-02	-3.507294e-03
## n_tokens_title	3.141392e-03	2.381394e-03
## n_tokens_content	-1.123327e-02	4.503348e-04
## n_unique_tokens	4.238245e-02	4.495002e-03
## n_non_stop_words	4.224852e-02	5.609304e-03
## n_non_stop_unique_tokens	3.652559e-02	6.804124e-03
## num_hrefs	2.290850e-02	-7.728561e-03
## num_self_hrefs	2.194085e-02	1.208016e-02
## num_imgs	1.794218e-02	-5.279523e-03
## num_videos	3.455053e-02	1.200655e-02
## average_token_length	3.551293e-02	1.104076e-03
## num_keywords	6.272493e-03	-2.148260e-02
## data_channel_is_lifestyle	-1.368446e-03	-8.784842e-03
## data_channel_is_entertainment	-2.414249e-02	3.129825e-02
## data_channel_is_bus	-8.042479e-03	1.828856e-02
## data_channel_is_socmed	1.820554e-02	-1.660382e-02
## data_channel_is_tech	2.268543e-02	2.544870e-03
## data_channel_is_world	-4.385291e-02	-1.172076e-02
## kw_min_min	-3.581221e-02	4.022713e-03
## kw_max_min	3.640957e-02	7.600646e-03
## kw_avg_min	3.100507e-02	2.929759e-03
## kw_min_max	2.043898e-02	-1.330452e-02
## kw_max_max	4.232766e-02	-1.038529e-02
## kw_avg_max	8.225331e-02	-2.909271e-03
## kw_min_avg	5.088991e-02	-1.475498e-02
## kw_max_avg	1.501448e-01	-8.343716e-03
## kw_avg_avg	1.534355e-01	-2.349314e-02
## self_reference_min_shares	8.198209e-01	-4.730205e-03
## self_reference_max_shares	8.582433e-01	-3.787376e-03
## self_reference_avg_shares	1.000000e+00	-3.470084e-03
## weekday_is_monday	-3.470084e-03	1.000000e+00
## weekday_is_tuesday	-3.714911e-03	-2.150149e-01
## weekday_is_wednesday	3.751128e-03	-2.162412e-01
## weekday_is_thursday	2.692735e-04	-2.135048e-01
## weekday_is_friday	3.410654e-03	-1.841283e-01
## weekday_is_saturday	-3.667544e-03	-1.133377e-01
## weekday_is_sunday	3.383416e-03	-1.195732e-01
## is_weekend	-8.730127e-05	-1.706096e-01
## LDA_00	1.984442e-03	5.715845e-03
## LDA_01	-1.881449e-02	2.558881e-02
## LDA_02	-4.396977e-02	-1.668046e-02
## LDA_03	4.355380e-02	-8.160272e-03
## LDA_04	1.308702e-02	2.036347e-05
## global_subjectivity	6.459935e-02	-7.431834e-03
## global_sentiment_polarity	6.003691e-03	-6.198921e-03

## global_rate_positive_words	8.641438e-03	-4.306792e-03
## global_rate_negative_words	1.891997e-02	1.064555e-02
## rate_positive_words	2.248835e-02	-2.338249e-03
## rate_negative_words	1.717888e-02	8.543686e-03
## avg_positive_polarity	3.781514e-02	5.424745e-03
## min_positive_polarity	1.699005e-02	-1.136379e-04
## max_positive_polarity	1.877462e-02	1.137356e-02
## avg_negative_polarity	-4.492938e-02	1.242630e-03
## min_negative_polarity	-2.111996e-02	8.555899e-04
## max_negative_polarity	-4.142404e-02	7.534642e-03
## title_subjectivity	6.737642e-03	-1.134895e-02
## title_sentiment_polarity	-5.545349e-04	-4.614218e-03
## abs_title_subjectivity	-1.086486e-03	-3.326161e-03
## abs_title_sentiment_polarity	6.244492e-03	-1.078695e-02
## shares	5.719484e-02	-2.269312e-02
##	weekday_is_tuesday	weekday_is_wednesday
## timedelta	-0.0029793827	0.0086005550
## n_tokens_title	0.0078714741	0.0106203192
## n_tokens_content	-0.0051868052	-0.0214339423
## n_unique_tokens	0.0032352228	0.0072990724
## n_non_stop_words	0.0020998442	-0.0033731522
## n_non_stop_unique_tokens	0.0022028628	0.0085737630
## num_hrefs	-0.0110835968	-0.0318249020
## num_self_hrefs	0.0015388713	-0.0189077636
## num_imgs	-0.0051657269	-0.0220196592
## num_videos	0.0066568121	-0.0016372867
## average_token_length	-0.0021450078	-0.0038342842
## num_keywords	-0.0093379414	-0.0189137480
## data_channel_is_lifestyle	-0.0158402269	-0.0013041698
## data_channel_is_entertainment	-0.0032272333	-0.0026542661
## data_channel_is_bus	0.0011865906	0.0175014155
## data_channel_is_socmed	0.0080187442	-0.0050552165
## data_channel_is_tech	0.0157707549	0.0027577167
## data_channel_is_world	-0.0048686518	-0.0014840778
## kw_min_min	-0.0067696695	0.0040104342
## kw_max_min	-0.0028911764	0.0007276445
## kw_avg_min	-0.0035470265	0.0007906250
## kw_min_max	0.0004519030	0.0090946148
## kw_max_max	0.0059012443	-0.0099909585
## kw_avg_max	0.0068859265	0.0048200175
## kw_min_avg	-0.0045049883	-0.0127289654
## kw_max_avg	-0.0068649856	-0.0044782944
## kw_avg_avg	-0.0060728009	-0.0133990764
## self_reference_min_shares	-0.0031929091	-0.0013576370
## self_reference_max_shares	-0.0024561016	0.0036295305
## self_reference_avg_shareess	-0.0037149108	0.0037511280
## weekday_is_monday	-0.2150149089	-0.2162411834
## weekday_is_tuesday	1.0000000000	-0.2318331507
## weekday_is_wednesday	-0.2318331507	1.0000000000
## weekday_is_thursday	-0.2288994415	-0.2302049025
## weekday_is_friday	-0.1974047897	-0.1985306301
## weekday_is_saturday	-0.1215098380	-0.1222028338
## weekday_is_sunday	-0.1281949537	-0.1289260761
## is_weekend	-0.1829112752	-0.1839544561
## LDA_00	-0.0000429849	0.0133243823
## LDA_01	-0.0107951842	-0.0090258709

## LDA_02	0.0003889473	-0.0001103001
## LDA_03	-0.0051096955	-0.0104278063
## LDA_04	0.0129517527	0.0051937147
## global_subjectivity	-0.0064663605	-0.0065931374
## global_sentiment_polarity	0.0058512777	0.0009015807
## global_rate_positive_words	0.0028157413	-0.0063980650
## global_rate_negative_words	-0.0134394548	-0.0173797644
## rate_positive_words	0.0150319045	0.0065130650
## rate_negative_words	-0.0164824770	-0.0110330486
## avg_positive_polarity	-0.0119999717	-0.0051077675
## min_positive_polarity	-0.0079494533	0.0018369362
## max_positive_polarity	-0.0065423261	-0.0117079123
## avg_negative_polarity	0.0115969529	0.0109740483
## min_negative_polarity	0.0130244682	0.0130726178
## max_negative_polarity	-0.0004775700	0.0020961939
## title_subjectivity	-0.0021203454	-0.0111899654
## title_sentiment_polarity	0.0008263177	-0.0147041589
## abs_title_subjectivity	0.0103376823	0.0093462953
## abs_title_sentiment_polarity	-0.0021774524	-0.0123821436
## shares	-0.0391807770	-0.0417559292
##	weekday_is_thursday	weekday_is_friday
## timedelta	0.0027366392	-3.494265e-03
## n_tokens_title	-0.0141776127	-1.737829e-03
## n_tokens_content	-0.0060066282	-1.452134e-02
## n_unique_tokens	0.0034508201	1.406121e-02
## n_non_stop_words	0.0016293416	6.250605e-04
## n_non_stop_unique_tokens	0.0027411523	1.312787e-02
## num_hrefs	-0.0133057208	-1.154536e-03
## num_self_hrefs	-0.0172978510	-2.244426e-02
## num_imgs	-0.0076521440	-8.035742e-03
## num_videos	-0.0035652389	-6.761799e-04
## average_token_length	0.0038347474	8.560885e-05
## num_keywords	-0.0116766354	1.433523e-03
## data_channel_is_lifestyle	-0.0089771922	2.573411e-03
## data_channel_is_entertainment	-0.0118999459	-9.006915e-03
## data_channel_is_bus	0.0178937872	-1.276545e-02
## data_channel_is_socmed	0.0100366684	-4.069481e-03
## data_channel_is_tech	-0.0039270259	-9.565629e-03
## data_channel_is_world	0.0014082204	1.840871e-02
## kw_min_min	0.0046636336	1.917593e-03
## kw_max_min	0.0005834359	2.616779e-04
## kw_avg_min	0.0008680735	3.608003e-03
## kw_min_max	0.0036309300	-3.234650e-03
## kw_max_max	-0.0051739833	3.569906e-03
## kw_avg_max	0.0093357983	3.380585e-03
## kw_min_avg	-0.0092367649	-1.437964e-03
## kw_max_avg	0.0004508515	6.307004e-04
## kw_avg_avg	-0.0041020941	4.679687e-03
## self_reference_min_shares	0.0040249689	1.691445e-03
## self_reference_max_shares	-0.0051789111	3.404696e-03
## self_reference_avg_shares	0.0002692735	3.410654e-03
## weekday_is_monday	-0.2135047812	-1.841283e-01
## weekday_is_tuesday	-0.2288994415	-1.974048e-01
## weekday_is_wednesday	-0.2302049025	-1.985306e-01
## weekday_is_thursday	1.0000000000	-1.960183e-01
## weekday_is_friday	-0.1960183443	1.000000e+00

## weekday_is_saturday	-0.1206564304	-1.040551e-01
## weekday_is_sunday	-0.1272945941	-1.097799e-01
## is_weekend	-0.1816266230	-1.566363e-01
## LDA_00	0.0223778174	-1.721792e-02
## LDA_01	-0.0100724605	-8.691201e-03
## LDA_02	-0.0002777138	2.003245e-02
## LDA_03	-0.0060882944	8.555331e-03
## LDA_04	-0.0064080364	-6.139224e-03
## global_subjectivity	-0.0032940070	8.347850e-03
## global_sentiment_polarity	-0.0018580584	-1.467108e-02
## global_rate_positive_words	-0.0024452202	-1.816654e-02
## global_rate_negative_words	0.0007445676	1.071690e-02
## rate_positive_words	0.0009295696	-1.626521e-02
## rate_negative_words	0.0009394684	2.002532e-02
## avg_positive_polarity	-0.0088372348	-3.960579e-04
## min_positive_polarity	0.0010999195	1.477259e-02
## max_positive_polarity	-0.0092649941	-1.478664e-02
## avg_negative_polarity	0.0067972184	-6.926059e-03
## min_negative_polarity	0.0058881833	2.078651e-03
## max_negative_polarity	-0.0005472139	-1.112698e-02
## title_subjectivity	0.0037129285	-6.468420e-04
## title_sentiment_polarity	-0.0049805253	-6.155796e-03
## abs_title_subjectivity	0.0038160113	1.135432e-02
## abs_title_sentiment_polarity	-0.0050110437	-2.889103e-03
## shares	-0.0250423928	9.257667e-03
##	weekday_is_saturday	weekday_is_sunday
## timedelta	-0.010298069	0.0068943950
## n_tokens_title	-0.012108842	0.0035327593
## n_tokens_content	0.038205364	0.0345137883
## n_unique_tokens	-0.038052427	-0.0121664317
## n_non_stop_words	-0.009215065	-0.0010106533
## n_non_stop_unique_tokens	-0.033875319	-0.0175719591
## num_hrefs	0.062326323	0.0418067134
## num_self_hrefs	0.050618057	0.0195431350
## num_imgs	0.035116999	0.0403366565
## num_videos	-0.005735244	-0.0138346263
## average_token_length	-0.005596734	0.0069728557
## num_keywords	0.041784230	0.0527869859
## data_channel_is_lifestyle	0.023781378	0.0276606130
## data_channel_is_entertainment	-0.014954659	0.0079076618
## data_channel_is_bus	-0.045005565	-0.0236520231
## data_channel_is_socmed	0.021900246	-0.0105610861
## data_channel_is_tech	0.018074582	-0.0304603280
## data_channel_is_world	0.000226631	-0.0008924224
## kw_min_min	-0.020078565	0.0074630182
## kw_max_min	-0.004221490	-0.0052426390
## kw_avg_min	-0.004965971	-0.0017793343
## kw_min_max	0.007688359	-0.0035150584
## kw_max_max	0.023479851	0.0026210114
## kw_avg_max	-0.006387125	-0.0272241930
## kw_min_avg	0.039037153	0.0283060814
## kw_max_avg	0.016497648	0.0129341380
## kw_avg_avg	0.036980218	0.0302187109
## self_reference_min_shares	-0.002821745	0.0082436411
## self_reference_max_shares	0.002784834	0.0044210649
## self_reference_avg_shares	-0.003667544	0.0033834160

## weekday_is_monday	-0.113337679	-0.1195731862	
## weekday_is_tuesday	-0.121509838	-0.1281949537	
## weekday_is_wednesday	-0.122202834	-0.1289260761	
## weekday_is_thursday	-0.120656430	-0.1272945941	
## weekday_is_friday	-0.104055113	-0.1097799209	
## weekday_is_saturday	1.000000000	-0.0675735397	
## weekday_is_sunday	-0.067573540	1.000000000	
## is_weekend	0.664310267	0.7008586735	
## LDA_00	-0.023066854	-0.0179896449	
## LDA_01	-0.001494524	0.0221717604	
## LDA_02	0.006039383	-0.0090692370	
## LDA_03	0.001800386	0.0322937126	
## LDA_04	0.014366189	-0.0235251759	
## global_subjectivity	0.014447004	0.0111717480	
## global_sentiment_polarity	0.012843122	0.0100433363	
## global_rate_positive_words	0.017174376	0.0251461470	
## global_rate_negative_words	0.005872022	0.0105245594	
## rate_positive_words	-0.003989643	-0.0049548351	
## rate_negative_words	-0.004687043	0.0051006220	
## avg_positive_polarity	0.013517355	0.0201245509	
## min_positive_polarity	-0.019523300	0.0058000371	
## max_positive_polarity	0.026896274	0.0212523995	
## avg_negative_polarity	-0.017703668	-0.0211818249	
## min_negative_polarity	-0.031707598	-0.0240601939	
## max_negative_polarity	0.007870506	-0.0047789491	
## title_subjectivity	0.008563267	0.0248389710	
## title_sentiment_polarity	0.024328653	0.0219457739	
## abs_title_subjectivity	-0.020577629	-0.0282217606	
## abs_title_sentiment_polarity	0.016149031	0.0354823620	
## shares	0.101764042	0.0897565402	
##	is_weekend	LDA_00	LDA_01
## timedelta	-2.197374e-03	0.0827905949	0.0018283017
## n_tokens_title	-6.010391e-03	-0.0711975205	0.0633442172
## n_tokens_content	5.317100e-02	0.0113384246	-0.0029827251
## n_unique_tokens	-3.631970e-02	0.0392700507	0.0549743888
## n_non_stop_words	-7.345317e-03	0.0745947479	0.0080584702
## n_non_stop_unique_tokens	-3.738300e-02	0.0833939186	0.0428717573
## num_hrefs	7.587953e-02	-0.0222964521	-0.0527646800
## num_self_hrefs	5.082970e-02	-0.0272265624	-0.0016481582
## num_imgs	5.532533e-02	-0.1477662824	0.0572390885
## num_videos	-1.446480e-02	-0.0739914713	0.0188465189
## average_token_length	1.222553e-03	0.0795446643	-0.0285182767
## num_keywords	6.941937e-02	-0.1587356221	-0.0629639382
## data_channel_is_lifestyle	3.772460e-02	-0.0085640756	-0.0790476691
## data_channel_is_entertainment	-4.767421e-03	-0.2125885414	0.6029099749
## data_channel_is_bus	-4.989539e-02	0.7844410589	-0.1293727477
## data_channel_is_socmed	7.745195e-03	0.1751659296	-0.0658547724
## data_channel_is_tech	-9.897804e-03	-0.2038757883	-0.1651343366
## data_channel_is_world	-5.065492e-04	-0.2453241492	-0.2139233094
## kw_min_min	-8.763790e-03	0.0299362958	0.0072783280
## kw_max_min	-6.945712e-03	-0.0055070296	-0.0031861212
## kw_avg_min	-4.883366e-03	0.0212850965	0.0003072777
## kw_min_max	2.863298e-03	0.0526944741	-0.0106903828
## kw_max_max	1.875013e-02	-0.0293138889	-0.0024778594
## kw_avg_max	-2.496190e-02	0.1288351547	-0.0766318711
## kw_min_avg	4.911505e-02	0.0041880270	-0.0087923884

## kw_max_avg	2.148459e-02	-0.0174753201	-0.0006440454
## kw_avg_avg	4.907736e-02	-0.0337448455	0.0154544313
## self_reference_min_shares	4.158521e-03	-0.0023963768	-0.0134731558
## self_reference_max_shares	5.303100e-03	0.0086112675	-0.0210786578
## self_reference_avg_sharess	-8.730127e-05	0.0019844417	-0.0188144919
## weekday_is_monday	-1.706096e-01	0.0057158454	0.0255888105
## weekday_is_tuesday	-1.829113e-01	-0.0000429849	-0.0107951842
## weekday_is_wednesday	-1.839545e-01	0.0133243823	-0.0090258709
## weekday_is_thursday	-1.816266e-01	0.0223778174	-0.0100724605
## weekday_is_friday	-1.566363e-01	-0.0172179177	-0.0086912009
## weekday_is_saturday	6.643103e-01	-0.0230668539	-0.0014945241
## weekday_is_sunday	7.008587e-01	-0.0179896449	0.0221717604
## is_weekend	1.000000e+00	-0.0299685748	0.0155419160
## LDA_00	-2.996857e-02	1.0000000000	-0.1782178529
## LDA_01	1.554192e-02	-0.1782178529	1.0000000000
## LDA_02	-2.476632e-03	-0.2695946390	-0.2391712266
## LDA_03	2.548062e-02	-0.2630301371	-0.1159179565
## LDA_04	-7.353459e-03	-0.2461775575	-0.2459022503
## global_subjectivity	1.869819e-02	-0.0059504637	0.0563104879
## global_sentiment_polarity	1.670615e-02	0.1110244841	-0.0197583744
## global_rate_positive_words	3.111727e-02	0.1414020936	0.0655197827
## global_rate_negative_words	1.208279e-02	-0.0829095860	0.1037769449
## rate_positive_words	-6.564343e-03	0.1623898554	-0.0194622173
## rate_negative_words	4.703019e-04	-0.1175774882	0.0322330303
## avg_positive_polarity	2.474072e-02	0.0109050182	0.0456056000
## min_positive_polarity	-9.612667e-03	-0.0961178716	0.0242186645
## max_positive_polarity	3.515071e-02	0.0533083027	0.0432257475
## avg_negative_polarity	-2.852573e-02	0.0442949728	-0.1066631274
## min_negative_polarity	-4.069401e-02	0.0502048642	-0.0674175853
## max_negative_polarity	2.046653e-03	-0.0124954344	-0.0333864300
## title_subjectivity	2.473077e-02	-0.0455692666	0.0586395709
## title_sentiment_polarity	3.383449e-02	0.0334557967	-0.0082569326
## abs_title_subjectivity	-3.585453e-02	-0.0020905625	-0.0727786039
## abs_title_sentiment_polarity	3.812779e-02	-0.0243306332	0.0382157669
## shares	1.399974e-01	0.0756263689	-0.0767499061
##	LDA_02	LDA_03	LDA_04
## timedelta	-1.492859e-01	-0.0235064895	9.523879e-02
## n_tokens_title	3.662239e-02	0.0430575342	-6.260302e-02
## n_tokens_content	9.185936e-02	-0.1309689876	3.046835e-02
## n_unique_tokens	-8.147576e-02	0.0037946292	-2.650426e-04
## n_non_stop_words	1.628846e-02	-0.1828162993	9.118466e-02
## n_non_stop_unique_tokens	-2.949046e-02	-0.1206123328	4.047637e-02
## num_hrefs	5.170904e-05	0.1154260379	-5.416315e-02
## num_self_hrefs	-8.894789e-02	-0.0250618535	1.396231e-01
## num_imgs	-8.870890e-02	0.1970160873	-1.613915e-02
## num_videos	-9.077214e-02	0.2724511483	-1.270315e-01
## average_token_length	1.071300e-01	-0.1877577418	2.891304e-02
## num_keywords	-1.528225e-02	0.0193298459	1.887556e-01
## data_channel_is_lifestyle	-1.186130e-01	-0.0582449260	2.441003e-01
## data_channel_is_entertainment	-2.178212e-01	0.2319348114	-2.787902e-01
## data_channel_is_bus	-2.216886e-01	-0.2255954454	-1.717510e-01
## data_channel_is_socmed	-2.080238e-02	-0.0226390578	-6.633991e-02
## data_channel_is_tech	-1.928610e-01	-0.2484479057	7.507103e-01
## data_channel_is_world	8.412514e-01	-0.2648245985	-1.886682e-01
## kw_min_min	-7.961381e-02	0.0032689350	4.325328e-02
## kw_max_min	-1.449277e-02	0.0190518628	2.970046e-03

## kw_avg_min	-3.333623e-02	0.0016371483	1.195339e-02
## kw_min_max	-5.623261e-02	0.0736896811	-5.705621e-02
## kw_max_max	7.812206e-02	0.0002643329	-4.950109e-02
## kw_avg_max	-1.123637e-01	0.2744378620	-2.196973e-01
## kw_min_avg	-1.166254e-01	0.1744846909	-5.419460e-02
## kw_max_avg	-9.922038e-02	0.1980507015	-8.140109e-02
## kw_avg_avg	-2.709731e-01	0.4328770857	-1.408983e-01
## self_reference_min_shares	-3.274662e-02	0.0301757343	1.511471e-02
## self_reference_max_shares	-4.212787e-02	0.0419246809	8.551555e-03
## self_reference_avg_shares	-4.396977e-02	0.0435537962	1.308702e-02
## weekday_is_monday	-1.668046e-02	-0.0081602722	2.036347e-05
## weekday_is_tuesday	3.889473e-04	-0.0051096955	1.295175e-02
## weekday_is_wednesday	-1.103001e-04	-0.0104278063	5.193715e-03
## weekday_is_thursday	-2.777138e-04	-0.0060882944	-6.408036e-03
## weekday_is_friday	2.003245e-02	0.0085553311	-6.139224e-03
## weekday_is_saturday	6.039383e-03	0.0018003863	1.436619e-02
## weekday_is_sunday	-9.069237e-03	0.0322937126	-2.352518e-02
## is_weekend	-2.476632e-03	0.0254806164	-7.353459e-03
## LDA_00	-2.695946e-01	-0.2630301371	-2.461776e-01
## LDA_01	-2.391712e-01	-0.1159179565	-2.459023e-01
## LDA_02	1.000000e+00	-0.3227854867	-2.479415e-01
## LDA_03	-3.227855e-01	1.0000000000	-3.433716e-01
## LDA_04	-2.479415e-01	-0.3433715722	1.000000e+00
## global_subjectivity	-1.805099e-01	0.0754320790	6.749932e-02
## global_sentiment_polarity	-2.296214e-01	-0.0010061003	1.436549e-01
## global_rate_positive_words	-2.404121e-01	-0.0377125927	9.819147e-02
## global_rate_negative_words	2.691319e-02	0.0701608212	-9.997289e-02
## rate_positive_words	-1.527708e-01	-0.1562641145	1.742074e-01
## rate_negative_words	2.012360e-01	-0.0043731863	-1.135516e-01
## avg_positive_polarity	-1.385631e-01	0.0658991561	2.794866e-02
## min_positive_polarity	-5.683608e-02	0.0976074331	2.889331e-02
## max_positive_polarity	-1.107355e-01	-0.0133496940	4.207418e-02
## avg_negative_polarity	3.112289e-02	-0.1098409930	1.189568e-01
## min_negative_polarity	-7.937074e-02	-0.0233390512	1.078683e-01
## max_negative_polarity	8.118080e-02	-0.0733697604	2.879998e-02
## title_subjectivity	-7.181503e-02	0.1216328051	-5.236253e-02
## title_sentiment_polarity	-7.484475e-02	0.0255632697	2.498381e-02
## abs_title_subjectivity	6.153059e-02	-0.0142374194	1.017631e-02
## abs_title_sentiment_polarity	-7.866749e-02	0.1166077667	-4.434845e-02
## shares	-1.366928e-01	0.0392756136	8.673353e-02
##	global_subjectivity global_sentiment_polarity		
## timedelta	0.126758465	0.1612219787	
## n_tokens_title	-0.053310634	-0.0679815713	
## n_tokens_content	0.130185642	0.0129594896	
## n_unique_tokens	0.464630813	0.1708459803	
## n_non_stop_words	0.651891189	0.2088270873	
## n_non_stop_unique_tokens	0.505801442	0.1701375879	
## num_hrefs	0.198600426	0.0886037598	
## num_self_hrefs	0.113749575	0.0919807964	
## num_imgs	0.076468984	0.0235957300	
## num_videos	0.080458692	-0.0324520654	
## average_token_length	0.576404687	0.1674535583	
## num_keywords	0.041607819	0.0870387629	
## data_channel_is_lifestyle	0.064573555	0.0827993582	
## data_channel_is_entertainment	0.038704789	-0.0375126143	
## data_channel_is_bus	-0.028384838	0.0758566798	



## data_channel_is_socmed	0.040181736	0.0731248855
## data_channel_is_tech	0.062245880	0.1361155206
## data_channel_is_world	-0.180816531	-0.2420297143
## kw_min_min	0.040822516	0.0777367771
## kw_max_min	0.013962455	0.0109083135
## kw_avg_min	0.027759014	0.0317358159
## kw_min_max	-0.017645619	-0.0126078166
## kw_max_max	-0.044019431	-0.0877603215
## kw_avg_max	-0.065534423	-0.0835624798
## kw_min_avg	0.008989125	-0.0010813659
## kw_max_avg	0.040883986	0.0132591478
## kw_avg_avg	0.063742734	0.0222001707
## self_reference_min_shares	0.049953064	0.0060529947
## self_reference_max_shares	0.060646061	0.0074170087
## self_reference_avg_shareess	0.064599345	0.0060036912
## weekday_is_monday	-0.007431834	-0.0061989214
## weekday_is_tuesday	-0.006466360	0.0058512777
## weekday_is_wednesday	-0.006593137	0.0009015807
## weekday_is_thursday	-0.003294007	-0.0018580584
## weekday_is_friday	0.008347850	-0.0146710759
## weekday_is_saturday	0.014447004	0.0128431219
## weekday_is_sunday	0.011171748	0.0100433363
## is_weekend	0.018698190	0.0167061476
## LDA_00	-0.005950464	0.1110244841
## LDA_01	0.056310488	-0.0197583744
## LDA_02	-0.180509866	-0.2296213674
## LDA_03	0.075432079	-0.0010061003
## LDA_04	0.067499317	0.1436549390
## global_subjectivity	1.000000000	0.3402244163
## global_sentiment_polarity	0.340224416	1.0000000000
## global_rate_positive_words	0.465604443	0.5752863313
## global_rate_negative_words	0.242831571	-0.4746391673
## rate_positive_words	0.477242060	0.7304779687
## rate_negative_words	0.111177279	-0.6595081378
## avg_positive_polarity	0.617721309	0.4989037494
## min_positive_polarity	0.229215821	0.0829171352
## max_positive_polarity	0.506233007	0.4303668849
## avg_negative_polarity	-0.431697477	0.2369277136
## min_negative_polarity	-0.340157817	0.2936816256
## max_negative_polarity	-0.213913463	-0.0422743952
## title_subjectivity	0.113219577	0.0248227903
## title_sentiment_polarity	0.035991476	0.2386800940
## abs_title_subjectivity	-0.006322016	-0.0362249681
## abs_title_sentiment_polarity	0.089820126	0.0696322339
## shares	0.058290455	0.0632611305
##	global_rate_positive_words	
## timedelta	0.2049534243	
## n_tokens_title	-0.0598916612	
## n_tokens_content	0.1287556595	
## n_unique_tokens	0.2814757038	
## n_non_stop_words	0.3850456166	
## n_non_stop_unique_tokens	0.3332967973	
## num_hrefs	0.0602935502	
## num_self_hrefs	0.1217489782	
## num_imgs	-0.0332051985	
## num_videos	0.0734856197	

## average_token_length	0.3060818259	
## num_keywords	0.0533543243	
## data_channel_is_lifestyle	0.0661453844	
## data_channel_is_entertainment	0.0290134218	
## data_channel_is_bus	0.0870344016	
## data_channel_is_socmed	0.0984379372	
## data_channel_is_tech	0.0927128546	
## data_channel_is_world	-0.2579145127	
## kw_min_min	0.0976574709	
## kw_max_min	0.0100678123	
## kw_avg_min	0.0368277185	
## kw_min_max	-0.0104696172	
## kw_max_max	-0.1079100698	
## kw_avg_max	-0.1058124934	
## kw_min_avg	-0.0052990783	
## kw_max_avg	0.0010372841	
## kw_avg_avg	-0.0005482842	
## self_reference_min_shares	0.0019413813	
## self_reference_max_shares	0.0183067553	
## self_reference_avg_shares	0.0086414382	
## weekday_is_monday	-0.0043067916	
## weekday_is_tuesday	0.0028157413	
## weekday_is_wednesday	-0.0063980650	
## weekday_is_thursday	-0.0024452202	
## weekday_is_friday	-0.0181665410	
## weekday_is_saturday	0.0171743758	
## weekday_is_sunday	0.0251461470	
## is_weekend	0.0311172719	
## LDA_00	0.1414020936	
## LDA_01	0.0655197827	
## LDA_02	-0.2404120905	
## LDA_03	-0.0377125927	
## LDA_04	0.0981914748	
## global_subjectivity	0.4656044433	
## global_sentiment_polarity	0.5752863313	
## global_rate_positive_words	1.0000000000	
## global_rate_negative_words	0.0997148607	
## rate_positive_words	0.6241936436	
## rate_negative_words	-0.3457093808	
## avg_positive_polarity	0.3272255082	
## min_positive_polarity	-0.1139786924	
## max_positive_polarity	0.4702986281	
## avg_negative_polarity	-0.1210013067	
## min_negative_polarity	-0.1031790455	
## max_negative_polarity	-0.0667732992	
## title_subjectivity	0.1149251795	
## title_sentiment_polarity	0.1405825188	
## abs_title_subjectivity	-0.1448994967	
## abs_title_sentiment_polarity	0.1069234459	
## shares	0.0632632867	
##	global_rate_negative_words	rate_positive_words
## timedelta	0.0023847653	0.1960345318
## n_tokens_title	0.0167335998	-0.0632136808
## n_tokens_content	0.1323259212	0.0863511045
## n_unique_tokens	0.1595003710	0.4337401639
## n_non_stop_words	0.2599830942	0.6132261845

## n_non_stop_unique_tokens	0.2062772678	0.4780256117
## num_hrefs	0.0296474175	0.1011171528
## num_self_hrefs	0.0095762196	0.1394031604
## num_imgs	0.0219303838	-0.0176314546
## num_videos	0.1857185419	-0.0469267424
## average_token_length	0.2182558069	0.5611271850
## num_keywords	-0.0423055704	0.0427800847
## data_channel_is_lifestyle	-0.0087310433	0.0538447597
## data_channel_is_entertainment	0.1108850476	-0.0419899240
## data_channel_is_bus	-0.0780983479	0.1264803995
## data_channel_is_socmed	-0.0209972388	0.0806175834
## data_channel_is_tech	-0.0979416250	0.1632488062
## data_channel_is_world	0.0258024368	-0.1716632396
## kw_min_min	-0.0032406195	0.0815873957
## kw_max_min	-0.0013263840	0.0074614217
## kw_avg_min	-0.0025277088	0.0317094945
## kw_min_max	0.0013907735	-0.0267523586
## kw_max_max	0.0077529910	-0.0893983023
## kw_avg_max	0.0033789888	-0.1451640017
## kw_min_avg	0.0169412988	-0.0564006042
## kw_max_avg	0.0132150395	-0.0339006680
## kw_avg_avg	0.0259119415	-0.0931143175
## self_reference_min_shares	0.0110310168	0.0188007962
## self_reference_max_shares	0.0200092280	0.0230599097
## self_reference_avg_sharess	0.0189199707	0.0224883546
## weekday_is_monday	0.0106455513	-0.0023382494
## weekday_is_tuesday	-0.0134394548	0.0150319045
## weekday_is_wednesday	-0.0173797644	0.0065130650
## weekday_is_thursday	0.0007445676	0.0009295696
## weekday_is_friday	0.0107169007	-0.0162652053
## weekday_is_saturday	0.0058720218	-0.0039896434
## weekday_is_sunday	0.0105245594	-0.0049548351
## is_weekend	0.0120827867	-0.0065643432
## LDA_00	-0.0829095860	0.1623898554
## LDA_01	0.1037769449	-0.0194622173
## LDA_02	0.0269131906	-0.1527708292
## LDA_03	0.0701608212	-0.1562641145
## LDA_04	-0.0999728901	0.1742074292
## global_subjectivity	0.2428315707	0.4772420599
## global_sentiment_polarity	-0.4746391673	0.7304779687
## global_rate_positive_words	0.0997148607	0.6241936436
## global_rate_negative_words	1.0000000000	-0.4186143973
## rate_positive_words	-0.4186143973	1.0000000000
## rate_negative_words	0.7782497716	-0.5576903736
## avg_positive_polarity	0.1843884195	0.4001288513
## min_positive_polarity	0.0540698023	0.0426397627
## max_positive_polarity	0.1736761235	0.4387552115
## avg_negative_polarity	-0.3483850230	-0.0180445162
## min_negative_polarity	-0.4773368411	0.1088213069
## max_negative_polarity	0.0770153860	-0.1946041718
## title_subjectivity	0.0888969560	-0.0194625227
## title_sentiment_polarity	-0.1363560336	0.1491677148
## abs_title_subjectivity	-0.0505223844	-0.0254485227
## abs_title_sentiment_polarity	0.0568251329	-0.0008198548
## shares	-0.0254607102	0.0447423476
##	rate_negative_words	avg_positive_polarity

## timedelta	-0.0822885533	0.1212115431
## n_tokens_title	0.0353420027	-0.0439326717
## n_tokens_content	0.1059174453	0.1330633428
## n_unique_tokens	0.1769621561	0.4091535850
## n_non_stop_words	0.3114487045	0.5803162333
## n_non_stop_unique_tokens	0.2185353744	0.4486984920
## num_hrefs	0.0526513961	0.1832534514
## num_self_hrefs	-0.0160004545	0.0949576139
## num_imgs	0.0343617682	0.0890817780
## num_videos	0.0738696730	0.0977056701
## average_token_length	0.3121483646	0.5215015566
## num_keywords	-0.0754916333	0.0337669405
## data_channel_is_lifestyle	-0.0351340633	0.0686245293
## data_channel_is_entertainment	0.0501451079	0.0556222560
## data_channel_is_bus	-0.0849968772	-0.0003380262
## data_channel_is_socmed	-0.0629950453	0.0175305613
## data_channel_is_tech	-0.1189665465	0.0190723891
## data_channel_is_world	0.2065499263	-0.1477883422
## kw_min_min	-0.0474713331	0.0445264974
## kw_max_min	-0.0084590926	0.0107667066
## kw_avg_min	-0.0192019399	0.0234948729
## kw_min_max	-0.0103569641	-0.0099704448
## kw_max_max	0.0529117278	-0.0478631076
## kw_avg_max	0.0003022880	-0.0692753147
## kw_min_avg	-0.0107983963	0.0085896859
## kw_max_avg	-0.0085407047	0.0298279689
## kw_avg_avg	-0.0374691238	0.0404843291
## self_reference_min_shares	0.0109612859	0.0277431542
## self_reference_max_shares	0.0150319901	0.0379032517
## self_reference_avg_share	0.0171788782	0.0378151410
## weekday_is_monday	0.0085436863	0.0054247447
## weekday_is_tuesday	-0.0164824770	-0.0119999717
## weekday_is_wednesday	-0.0110330486	-0.0051077675
## weekday_is_thursday	0.0009394684	-0.0088372348
## weekday_is_friday	0.0200253245	-0.0003960579
## weekday_is_saturday	-0.0046870428	0.0135173548
## weekday_is_sunday	0.0051006220	0.0201245509
## is_weekend	0.0004703019	0.0247407165
## LDA_00	-0.1175774882	0.0109050182
## LDA_01	0.0322330303	0.0456056000
## LDA_02	0.2012359867	-0.1385631044
## LDA_03	-0.0043731863	0.0658991561
## LDA_04	-0.1135515985	0.0279486613
## global_subjectivity	0.1111772791	0.6177213094
## global_sentiment_polarity	-0.6595081378	0.4989037494
## global_rate_positive_words	-0.3457093808	0.3272255082
## global_rate_negative_words	0.7782497716	0.1843884195
## rate_positive_words	-0.5576903736	0.4001288513
## rate_negative_words	1.0000000000	0.1300174395
## avg_positive_polarity	0.1300174395	1.0000000000
## min_positive_polarity	0.1883146600	0.4478392499
## max_positive_polarity	0.0186111801	0.6953845758
## avg_negative_polarity	-0.3413624451	-0.2624716920
## min_negative_polarity	-0.4500120785	-0.2204414125
## max_negative_polarity	0.0322701053	-0.1311969073
## title_subjectivity	-0.0072039062	0.0473872083

## title_sentiment_polarity	-0.1848778701	0.0819707807
## abs_title_subjectivity	0.0488153392	0.0159046620
## abs_title_sentiment_polarity	-0.0268572025	0.0957580667
## shares	-0.0675732295	0.0190065049
##	min_positive_polarity	max_positive_polarity
## timedelta	0.0533339006	0.091992413
## n_tokens_title	-0.0251261470	-0.013711803
## n_tokens_content	-0.2652683292	0.417965112
## n_unique_tokens	0.4067620461	0.138505909
## n_non_stop_words	0.2273009637	0.518774066
## n_non_stop_unique_tokens	0.3204868050	0.257142494
## num_hrefs	-0.0903684580	0.286218816
## num_self_hrefs	-0.0736776606	0.186815722
## num_imgs	-0.0344629463	0.158222073
## num_videos	-0.0139299326	0.124784885
## average_token_length	0.2114861766	0.451383245
## num_keywords	-0.0161886066	0.063047522
## data_channel_is_lifestyle	-0.0030672109	0.068905301
## data_channel_is_entertainment	-0.0163018590	0.081185938
## data_channel_is_bus	-0.0493607733	0.019503588
## data_channel_is_socmed	-0.0491851901	0.028126015
## data_channel_is_tech	0.0303018056	0.028763303
## data_channel_is_world	-0.0498410041	-0.123943583
## kw_min_min	0.0230583853	0.021482027
## kw_max_min	0.0044944052	0.013023707
## kw_avg_min	0.0053577922	0.025024657
## kw_min_max	0.0012941374	-0.020517210
## kw_max_max	-0.0249162804	-0.021206869
## kw_avg_max	0.0062409249	-0.097713160
## kw_min_avg	0.0199308803	-0.015224289
## kw_max_avg	0.0197383489	0.009300286
## kw_avg_avg	0.0484266071	-0.006844676
## self_reference_min_shares	0.0255105304	0.002522100
## self_reference_max_shares	-0.0023987496	0.037143277
## self_reference_avg_sharess	0.0169900472	0.018774617
## weekday_is_monday	-0.0001136379	0.011373558
## weekday_is_tuesday	-0.0079494533	-0.006542326
## weekday_is_wednesday	0.0018369362	-0.011707912
## weekday_is_thursday	0.0010999195	-0.009264994
## weekday_is_friday	0.0147725879	-0.014786639
## weekday_is_saturday	-0.0195233000	0.026896274
## weekday_is_sunday	0.0058000371	0.021252399
## is_weekend	-0.0096126668	0.035150714
## LDA_00	-0.0961178716	0.053308303
## LDA_01	0.0242186645	0.043225747
## LDA_02	-0.0568360841	-0.110735502
## LDA_03	0.0976074331	-0.013349694
## LDA_04	0.0288933058	0.042074180
## global_subjectivity	0.2292158210	0.506233007
## global_sentiment_polarity	0.0829171352	0.430366885
## global_rate_positive_words	-0.1139786924	0.470298628
## global_rate_negative_words	0.0540698023	0.173676123
## rate_positive_words	0.0426397627	0.438755211
## rate_negative_words	0.1883146600	0.018611180
## avg_positive_polarity	0.4478392499	0.695384576
## min_positive_polarity	1.0000000000	-0.005683202

## max_positive_polarity	-0.0056832017	1.000000000
## avg_negative_polarity	-0.0516019376	-0.259177498
## min_negative_polarity	0.1007576373	-0.352004360
## max_negative_polarity	-0.1790466384	0.002565641
## title_subjectivity	0.0047728800	0.051291814
## title_sentiment_polarity	-0.0006490695	0.078900414
## abs_title_subjectivity	0.0158872789	-0.018060764
## abs_title_sentiment_polarity	0.0081614742	0.081099568
## shares	-0.0320381275	0.033220447
##	avg_negative_polarity	min_negative_polarity
## timedelta	0.007344119	0.0723377665
## n_tokens_title	-0.018506115	-0.0347130659
## n_tokens_content	-0.137432282	-0.4567120829
## n_unique_tokens	-0.204363616	0.0438863610
## n_non_stop_words	-0.344362115	-0.3025074331
## n_non_stop_unique_tokens	-0.243354934	-0.0822209387
## num_hrefs	-0.148775061	-0.2609269679
## num_self_hrefs	-0.055201044	-0.1175609041
## num_imgs	-0.067511758	-0.1353264489
## num_videos	-0.118924312	-0.1417098512
## average_token_length	-0.310572073	-0.2660084318
## num_keywords	0.023460473	0.0006686681
## data_channel_is_lifestyle	-0.007507156	-0.0223341438
## data_channel_is_entertainment	-0.134033184	-0.1115385523
## data_channel_is_bus	0.060821432	0.0704959576
## data_channel_is_socmed	0.006036885	0.0070208990
## data_channel_is_tech	0.114139446	0.1191116895
## data_channel_is_world	0.021837859	-0.0802036440
## kw_min_min	0.016757468	0.0511376597
## kw_max_min	-0.006519384	-0.0032133776
## kw_avg_min	-0.004267932	0.0032159007
## kw_min_max	-0.004676647	0.0018909466
## kw_max_max	-0.018186639	-0.0560416289
## kw_avg_max	-0.018997532	0.0136443252
## kw_min_avg	-0.019233244	-0.0006702353
## kw_max_avg	-0.045455371	-0.0230615865
## kw_avg_avg	-0.065791975	-0.0132621296
## self_reference_min_shares	-0.030629838	-0.0047810671
## self_reference_max_shares	-0.047057792	-0.0402249292
## self_reference_avg_share	-0.044929375	-0.0211199622
## weekday_is_monday	0.001242630	0.0008555899
## weekday_is_tuesday	0.011596953	0.0130244682
## weekday_is_wednesday	0.010974048	0.0130726178
## weekday_is_thursday	0.006797218	0.0058881833
## weekday_is_friday	-0.006926059	0.0020786514
## weekday_is_saturday	-0.017703668	-0.0317075983
## weekday_is_sunday	-0.021181825	-0.0240601939
## is_weekend	-0.028525733	-0.0406940082
## LDA_00	0.044294973	0.0502048642
## LDA_01	-0.106663127	-0.0674175853
## LDA_02	0.031122893	-0.0793707388
## LDA_03	-0.109840993	-0.0233390512
## LDA_04	0.118956827	0.1078682980
## global_subjectivity	-0.431697477	-0.3401578167
## global_sentiment_polarity	0.236927714	0.2936816256
## global_rate_positive_words	-0.121001307	-0.1031790455

## global_rate_negative_words	-0.348385023	-0.4773368411
## rate_positive_words	-0.018044516	0.1088213069
## rate_negative_words	-0.341362445	-0.4500120785
## avg_positive_polarity	-0.262471692	-0.2204414125
## min_positive_polarity	-0.051601938	0.1007576373
## max_positive_polarity	-0.259177498	-0.3520043600
## avg_negative_polarity	1.000000000	0.7484337136
## min_negative_polarity	0.748433714	1.0000000000
## max_negative_polarity	0.573856044	0.0732807318
## title_subjectivity	-0.074181779	-0.0555101665
## title_sentiment_polarity	0.087377171	0.0809906004
## abs_title_subjectivity	-0.010837940	-0.0156393711
## abs_title_sentiment_polarity	-0.075029186	-0.0583644605
## shares	-0.003511512	-0.0048992452
##	max_negative_polarity	title_subjectivity
## timedelta	-0.0619027531	-0.017477311
## n_tokens_title	0.0143056445	0.084982103
## n_tokens_content	0.2272498497	0.003778906
## n_unique_tokens	-0.3281483679	-0.006014666
## n_non_stop_words	-0.1916611838	-0.029780988
## n_non_stop_unique_tokens	-0.2590593871	-0.027689561
## num_hrefs	0.0584170918	0.037901680
## num_self_hrefs	0.0403527772	-0.011507606
## num_imgs	0.0473107796	0.044339631
## num_videos	0.0297290417	0.063294952
## average_token_length	-0.1869904549	-0.041050062
## num_keywords	0.0270013290	0.014298651
## data_channel_is_lifestyle	0.0088917601	0.004311771
## data_channel_is_entertainment	-0.0140874536	0.054238918
## data_channel_is_bus	-0.0097482046	-0.042139107
## data_channel_is_socmed	-0.0131019197	-0.014969652
## data_channel_is_tech	0.0171474856	-0.041053663
## data_channel_is_world	0.0754476879	-0.060683504
## kw_min_min	-0.0279425828	-0.003169693
## kw_max_min	-0.0016062640	0.011839323
## kw_avg_min	-0.0072695900	0.007277047
## kw_min_max	-0.0066928923	0.020525743
## kw_max_max	0.0295382422	0.004181476
## kw_avg_max	-0.0215961464	0.044194274
## kw_min_avg	-0.0112862262	0.041965585
## kw_max_avg	-0.0221019680	0.043389911
## kw_avg_avg	-0.0446349077	0.088587070
## self_reference_min_shares	-0.0431243958	0.002078068
## self_reference_max_shares	-0.0217580558	0.010066119
## self_reference_avg_shares	-0.0414240405	0.006737642
## weekday_is_monday	0.0075346423	-0.011348946
## weekday_is_tuesday	-0.0004775700	-0.002120345
## weekday_is_wednesday	0.0020961939	-0.011189965
## weekday_is_thursday	-0.0005472139	0.003712928
## weekday_is_friday	-0.0111269824	-0.000646842
## weekday_is_saturday	0.0078705063	0.008563267
## weekday_is_sunday	-0.0047789491	0.024838971
## is_weekend	0.0020466528	0.024730768
## LDA_00	-0.0124954344	-0.045569267
## LDA_01	-0.0333864300	0.058639571
## LDA_02	0.0811807963	-0.071815025

## LDA_03	-0.0733697604	0.121632805
## LDA_04	0.0287999792	-0.052362530
## global_subjectivity	-0.2139134629	0.113219577
## global_sentiment_polarity	-0.0422743952	0.024822790
## global_rate_positive_words	-0.0667732992	0.114925180
## global_rate_negative_words	0.0770153860	0.088896956
## rate_positive_words	-0.1946041718	-0.019462523
## rate_negative_words	0.0322701053	-0.007203906
## avg_positive_polarity	-0.1311969073	0.047387208
## min_positive_polarity	-0.1790466384	0.004772880
## max_positive_polarity	0.0025656410	0.051291814
## avg_negative_polarity	0.5738560440	-0.074181779
## min_negative_polarity	0.0732807318	-0.055510166
## max_negative_polarity	1.0000000000	-0.017282647
## title_subjectivity	-0.0172826465	1.0000000000
## title_sentiment_polarity	-0.0006841802	0.228639393
## abs_title_subjectivity	-0.0009656253	-0.501053389
## abs_title_sentiment_polarity	-0.0174455186	0.716235962
## shares	0.0031002814	0.025858813
##	title_sentiment_polarity	abs_title_subjectivity
## timedelta	0.0399822493	0.0088593215
## n_tokens_title	0.0052808196	-0.1492604575
## n_tokens_content	0.0197520513	0.0109036178
## n_unique_tokens	-0.0222721666	-0.0071953430
## n_non_stop_words	-0.0051729722	0.0178826252
## n_non_stop_unique_tokens	-0.0308292011	0.0090435392
## num_hrefs	0.0352661935	0.0092121412
## num_self_hrefs	0.0248993025	0.0086652061
## num_imgs	0.0447434528	-0.0096679951
## num_videos	0.0163833176	-0.0263412425
## average_token_length	-0.0129598705	0.0267196129
## num_keywords	0.0292319340	-0.0095351693
## data_channel_is_lifestyle	0.0392779367	0.0057686484
## data_channel_is_entertainment	-0.0093056757	-0.0498677520
## data_channel_is_bus	0.0190566846	-0.0049787787
## data_channel_is_socmed	0.0260659006	0.0118792537
## data_channel_is_tech	0.0267629900	0.0071230743
## data_channel_is_world	-0.0853572063	0.0564847711
## kw_min_min	0.0320882097	0.0065097455
## kw_max_min	0.0152019338	-0.0030479510
## kw_avg_min	0.0171955749	0.0019349306
## kw_min_max	0.0090981265	-0.0100669266
## kw_max_max	-0.0296184875	-0.0080034279
## kw_avg_max	-0.0119624939	-0.0136027570
## kw_min_avg	0.0264624061	-0.0151542088
## kw_max_avg	0.0210335125	-0.0130944317
## kw_avg_avg	0.0365002622	-0.0242812137
## self_reference_min_shares	-0.0007524291	0.0032291581
## self_reference_max_shares	0.0021227431	-0.0038842205
## self_reference_avg_sharess	-0.0005545349	-0.0010864860
## weekday_is_monday	-0.0046142178	-0.0033261610
## weekday_is_tuesday	0.0008263177	0.0103376823
## weekday_is_wednesday	-0.0147041589	0.0093462953
## weekday_is_thursday	-0.0049805253	0.0038160113
## weekday_is_friday	-0.0061557965	0.0113543188
## weekday_is_saturday	0.0243286529	-0.0205776289



## weekday_is_sunday	0.0219457739	-0.0282217606
## is_weekend	0.0338344886	-0.0358545295
## LDA_00	0.0334557967	-0.0020905625
## LDA_01	-0.0082569326	-0.0727786039
## LDA_02	-0.0748447525	0.0615305871
## LDA_03	0.0255632697	-0.0142374194
## LDA_04	0.0249838135	0.0101763114
## global_subjectivity	0.0359914762	-0.0063220164
## global_sentiment_polarity	0.2386800940	-0.0362249681
## global_rate_positive_words	0.1405825188	-0.1448994967
## global_rate_negative_words	-0.1363560336	-0.0505223844
## rate_positive_words	0.1491677148	-0.0254485227
## rate_negative_words	-0.1848778701	0.0488153392
## avg_positive_polarity	0.0819707807	0.0159046620
## min_positive_polarity	-0.0006490695	0.0158872789
## max_positive_polarity	0.0789004136	-0.0180607636
## avg_negative_polarity	0.0873771712	-0.0108379396
## min_negative_polarity	0.0809906004	-0.0156393711
## max_negative_polarity	-0.0006841802	-0.0009656253
## title_subjectivity	0.2286393934	-0.5010533885
## title_sentiment_polarity	1.0000000000	-0.2460916738
## abs_title_subjectivity	-0.2460916738	1.0000000000
## abs_title_sentiment_polarity	0.4021626338	-0.4126119670
## shares	0.0452892036	0.0048310338
##	abs_title_sentiment_polarity	shares
## timedelta	-0.0018525326	0.036571726
## n_tokens_title	0.0487523088	-0.042049834
## n_tokens_content	0.0114885660	0.047820741
## n_unique_tokens	-0.0180250710	-0.049099714
## n_non_stop_words	-0.0269930140	-0.013187554
## n_non_stop_unique_tokens	-0.0368586463	-0.050807230
## num_hrefs	0.0502650280	0.077652396
## num_self_hrefs	-0.0104585510	0.043162084
## num_imgs	0.0544530836	0.055926834
## num_videos	0.0539458804	-0.002898373
## average_token_length	-0.0364047815	-0.025551830
## num_keywords	0.0195806869	0.065535166
## data_channel_is_lifestyle	0.0211357290	0.031436922
## data_channel_is_entertainment	0.0386364693	-0.105421752
## data_channel_is_bus	-0.0291278490	0.001639743
## data_channel_is_socmed	-0.0029163465	0.114944422
## data_channel_is_tech	-0.0423646232	0.097379147
## data_channel_is_world	-0.0643434945	-0.137430983
## kw_min_min	0.0094046515	0.039892832
## kw_max_min	0.0158622739	0.022471750
## kw_avg_min	0.0118642423	0.031629207
## kw_min_max	0.0219397266	0.007840949
## kw_max_max	-0.0069176430	-0.024916387
## kw_avg_max	0.0323038897	0.016024753
## kw_min_avg	0.0536614354	0.089510210
## kw_max_avg	0.0469152925	0.063157450
## kw_avg_avg	0.0933080525	0.147677561
## self_reference_min_shares	0.0041424972	0.044587709
## self_reference_max_shares	0.0071453441	0.054805408
## self_reference_avg_shares	0.0062444922	0.057194838
## weekday_is_monday	-0.0107869497	-0.022693122

## weekday_is_tuesday	-0.0021774524	-0.039180777
## weekday_is_wednesday	-0.0123821436	-0.041755929
## weekday_is_thursday	-0.0050110437	-0.025042393
## weekday_is_friday	-0.0028891025	0.009257667
## weekday_is_saturday	0.0161490305	0.101764042
## weekday_is_sunday	0.0354823620	0.089756540
## is_weekend	0.0381277925	0.139997449
## LDA_00	-0.0243306332	0.075626369
## LDA_01	0.0382157669	-0.076749906
## LDA_02	-0.0786674921	-0.136692819
## LDA_03	0.1166077667	0.039275614
## LDA_04	-0.0443484506	0.086733535
## global_subjectivity	0.0898201256	0.058290455
## global_sentiment_polarity	0.0696322339	0.063261131
## global_rate_positive_words	0.1069234459	0.063263287
## global_rate_negative_words	0.0568251329	-0.025460710
## rate_positive_words	-0.0008198548	0.044742348
## rate_negative_words	-0.0268572025	-0.067573229
## avg_positive_polarity	0.0957580667	0.019006505
## min_positive_polarity	0.0081614742	-0.032038128
## max_positive_polarity	0.0810995678	0.033220447
## avg_negative_polarity	-0.0750291863	-0.003511512
## min_negative_polarity	-0.0583644605	-0.004899245
## max_negative_polarity	-0.0174455186	0.003100281
## title_subjectivity	0.7162359621	0.025858813
## title_sentiment_polarity	0.4021626338	0.045289204
## abs_title_subjectivity	-0.4126119670	0.004831034
## abs_title_sentiment_polarity	1.0000000000	0.029515150
## shares	0.0295151503	1.000000000

```
# Showing descriptives  
summary(data_no_outliers)
```

```

##      url                timedelta      n_tokens_title n_tokens_content
## Length:35103          Min.   : 8.0    Min.   : 2.0    Min.   : 0.0
## Class :character      1st Qu.:162.0    1st Qu.: 9.0    1st Qu.: 250.0
## Mode  :character      Median :337.0    Median :10.0   Median : 414.0
##                               Mean  :353.7    Mean  :10.4   Mean  : 546.8
##                               3rd Qu.:545.0    3rd Qu.:12.0   3rd Qu.: 715.0
##                               Max.   :731.0    Max.   :20.0   Max.   :7764.0
## n_unique_tokens  n_non_stop_words n_non_stop_unique_tokens  num_hrefs
## Min.   :0.0000    Min.   :0.0000    Min.   :0.0000          Min.   : 0.00
## 1st Qu.:0.4718    1st Qu.:1.0000    1st Qu.:0.6276          1st Qu.: 4.00
## Median :0.5387    Median :1.0000    Median :0.6912          Median : 7.00
## Mean   :0.5313    Mean   :0.9724    Mean   :0.6748          Mean   :10.58
## 3rd Qu.:0.6070    3rd Qu.:1.0000    3rd Qu.:0.7545          3rd Qu.:13.00
## Max.   :1.0000    Max.   :1.0000    Max.   :1.0000          Max.   :187.00
## num_self_hrefs    num_imgs          num_videos    average_token_length
## Min.   : 0.000    Min.   : 0.000    Min.   : 0.000    Min.   :0.000
## 1st Qu.: 1.000    1st Qu.: 1.000    1st Qu.: 0.000    1st Qu.:4.482
## Median : 3.000    Median : 1.000    Median : 0.000    Median :4.668
## Mean   : 3.297    Mean   : 4.336    Mean   : 1.188    Mean   :4.560
## 3rd Qu.: 4.000    3rd Qu.: 3.000    3rd Qu.: 1.000    3rd Qu.:4.859
## Max.   :116.000    Max.   :128.000    Max.   :75.000    Max.   :8.042
## num_keywords      data_channel_is_lifestyle data_channel_is_entertainment
## Min.   : 1.000    Min.   :0.00000    Min.   :0.0000
## 1st Qu.: 6.000    1st Qu.:0.00000    1st Qu.:0.0000
## Median : 7.000    Median :0.00000    Median :0.0000
## Mean   : 7.199    Mean   :0.05122    Mean   :0.1808
## 3rd Qu.: 9.000    3rd Qu.:0.00000    3rd Qu.:0.0000
## Max.   :10.000    Max.   :1.00000    Max.   :1.0000
## data_channel_is_bus data_channel_is_socmed data_channel_is_tech
## Min.   :0.0000    Min.   :0.00000    Min.   :0.0000
## 1st Qu.:0.0000    1st Qu.:0.00000    1st Qu.:0.0000
## Median :0.0000    Median :0.00000    Median :0.0000
## Mean   :0.1638    Mean   :0.05666    Mean   :0.1866
## 3rd Qu.:0.0000    3rd Qu.:0.00000    3rd Qu.:0.0000
## Max.   :1.0000    Max.   :1.00000    Max.   :1.0000
## data_channel_is_world kw_min_min      kw_max_min      kw_avg_min
## Min.   :0.0000    Min.   : -1.00    Min.   : 0      Min.   : -1.0
## 1st Qu.:0.0000    1st Qu.: -1.00    1st Qu.: 444    1st Qu.: 140.7
## Median :0.0000    Median : -1.00    Median : 654    Median : 234.3
## Mean   :0.2244    Mean   : 26.23    Mean   : 1122    Mean   : 306.7
## 3rd Qu.:0.0000    3rd Qu.: 4.00    3rd Qu.: 1000    3rd Qu.: 355.3
## Max.   :1.0000    Max.   :377.00    Max.   :298400    Max.   :39979.0
## kw_min_max      kw_max_max      kw_avg_max      kw_min_avg
## Min.   : 0      Min.   : 0      Min.   : 0      Min.   : -1
## 1st Qu.: 0      1st Qu.:843300    1st Qu.:171816    1st Qu.: 0
## Median : 1400    Median :843300    Median :241825    Median :1002
## Mean   : 13355    Mean   :751073    Mean   :256558    Mean   :1088
## 3rd Qu.: 7400    3rd Qu.:843300    3rd Qu.:327462    3rd Qu.:1992
## Max.   :843300    Max.   :843300    Max.   :843300    Max.   :3610
## kw_max_avg      kw_avg_avg      self_reference_min_shares
## Min.   : 0      Min.   : 0      Min.   : 0
## 1st Qu.: 3531    1st Qu.: 2351    1st Qu.: 622
## Median : 4230    Median : 2816    Median : 1200
## Mean   : 5460    Mean   : 3055    Mean   : 3610
## 3rd Qu.: 5852    3rd Qu.: 3490    3rd Qu.: 2400

```

```

## Max.      :298400    Max.      :37608    Max.      :843300
## self_reference_max_shares self_reference_avg_shares weekday_is_monday
## Min.       :      0          Min.       :      0.0          Min.       :0.0000
## 1st Qu.: 1000          1st Qu.: 952.2          1st Qu.:0.0000
## Median : 2700          Median : 2100.0          Median :0.0000
## Mean      : 9563          Mean      : 5861.8          Mean      :0.1671
## 3rd Qu.: 7400          3rd Qu.: 4754.0          3rd Qu.:0.0000
## Max.      :843300          Max.      :843300.0          Max.      :1.0000
## weekday_is_tuesday weekday_is_wednesday weekday_is_thursday weekday_is_friday
## Min.       :0.0000    Min.       :0.0000    Min.       :0.0000    Min.       :0.0000
## 1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.0000    1st Qu.:0.0000
## Median :0.0000    Median :0.0000    Median :0.0000    Median :0.0000
## Mean      :0.1873    Mean      :0.1891    Mean      :0.1852    Mean      :0.1446
## 3rd Qu.:0.0000    3rd Qu.:0.0000    3rd Qu.:0.0000    3rd Qu.:0.0000
## Max.      :1.0000    Max.      :1.0000    Max.      :1.0000    Max.      :1.0000
## weekday_is_saturday weekday_is_sunday is_weekend LDA_00
## Min.       :0.00000    Min.       :0.00000    Min.       :0.0000    Min.       :0.01818
## 1st Qu.:0.00000    1st Qu.:0.00000    1st Qu.:0.0000    1st Qu.:0.02506
## Median :0.00000    Median :0.00000    Median :0.0000    Median :0.03341
## Mean      :0.06019    Mean      :0.06655    Mean      :0.1267    Mean      :0.18630
## 3rd Qu.:0.00000    3rd Qu.:0.00000    3rd Qu.:0.0000    3rd Qu.:0.24487
## Max.      :1.00000    Max.      :1.00000    Max.      :1.0000    Max.      :0.92699
## LDA_01 LDA_02 LDA_03 LDA_04
## Min.       :0.01818    Min.       :0.01818    Min.       :0.01818    Min.       :0.01818
## 1st Qu.:0.02501    1st Qu.:0.02857    1st Qu.:0.02553    1st Qu.:0.02857
## Median :0.03334    Median :0.04004    Median :0.04000    Median :0.05000
## Mean      :0.14162    Mean      :0.22611    Mean      :0.21044    Mean      :0.23553
## 3rd Qu.:0.15030    3rd Qu.:0.36594    3rd Qu.:0.31880    3rd Qu.:0.40317
## Max.      :0.92595    Max.      :0.92000    Max.      :0.92653    Max.      :0.92712
## global_subjectivity global_sentiment_polarity global_rate_positive_words
## Min.       :0.0000    Min.       : -0.39375          Min.       :0.00000
## 1st Qu.:0.3939    1st Qu.: 0.05762          1st Qu.:0.02832
## Median :0.4509    Median : 0.11855          Median :0.03892
## Mean      :0.4414    Mean      : 0.11890          Mean      :0.03958
## 3rd Qu.:0.5053    3rd Qu.: 0.17714          3rd Qu.:0.05025
## Max.      :1.0000    Max.      : 0.72784          Max.      :0.15549
## global_rate_negative_words rate_positive_words rate_negative_words
## Min.       :0.000000    Min.       :0.0000    Min.       :0.0000
## 1st Qu.:0.009595    1st Qu.:0.6000    1st Qu.:0.1859
## Median :0.015306    Median :0.7105    Median :0.2800
## Mean      :0.016562    Mean      :0.6838    Mean      :0.2884
## 3rd Qu.:0.021692    3rd Qu.:0.8000    3rd Qu.:0.3846
## Max.      :0.162037    Max.      :1.0000    Max.      :1.0000
## avg_positive_polarity min_positive_polarity max_positive_polarity
## Min.       :0.0000    Min.       :0.00000    Min.       :0.0000
## 1st Qu.:0.3056    1st Qu.:0.05000    1st Qu.:0.6000
## Median :0.3574    Median :0.10000    Median :0.8000
## Mean      :0.3529    Mean      :0.09521    Mean      :0.7551
## 3rd Qu.:0.4093    3rd Qu.:0.10000    3rd Qu.:1.0000
## Max.      :1.0000    Max.      :1.00000    Max.      :1.0000
## avg_negative_polarity min_negative_polarity max_negative_polarity
## Min.       : -1.0000    Min.       : -1.0000    Min.       : -1.0000
## 1st Qu.: -0.3255    1st Qu.: -0.7000    1st Qu.: -0.1250
## Median : -0.2512    Median : -0.5000    Median : -0.1000
## Mean      : -0.2578    Mean      : -0.5193    Mean      : -0.1069
## 3rd Qu.: -0.1855    3rd Qu.: -0.3000    3rd Qu.: -0.0500

```

```
## Max. : 0.0000      Max. : 0.0000      Max. : 0.0000
## title_subjectivity title_sentiment_polarity abs_title_subjectivity
## Min. :0.0000      Min. : -1.00000      Min. :0.0000
## 1st Qu.:0.0000      1st Qu.: 0.00000      1st Qu.:0.1667
## Median :0.1250      Median : 0.00000      Median :0.5000
## Mean :0.2769      Mean : 0.06861      Mean :0.3420
## 3rd Qu.:0.5000      3rd Qu.: 0.13636      3rd Qu.:0.5000
## Max. :1.0000      Max. : 1.00000      Max. :0.5000
## abs_title_sentiment_polarity shares
## Min. :0.000      Min. : 1
## 1st Qu.:0.000      1st Qu.: 903
## Median :0.000      Median :1300
## Mean :0.152      Mean :1672
## 3rd Qu.:0.250      3rd Qu.:2100
## Max. :1.000      Max. :5500
```

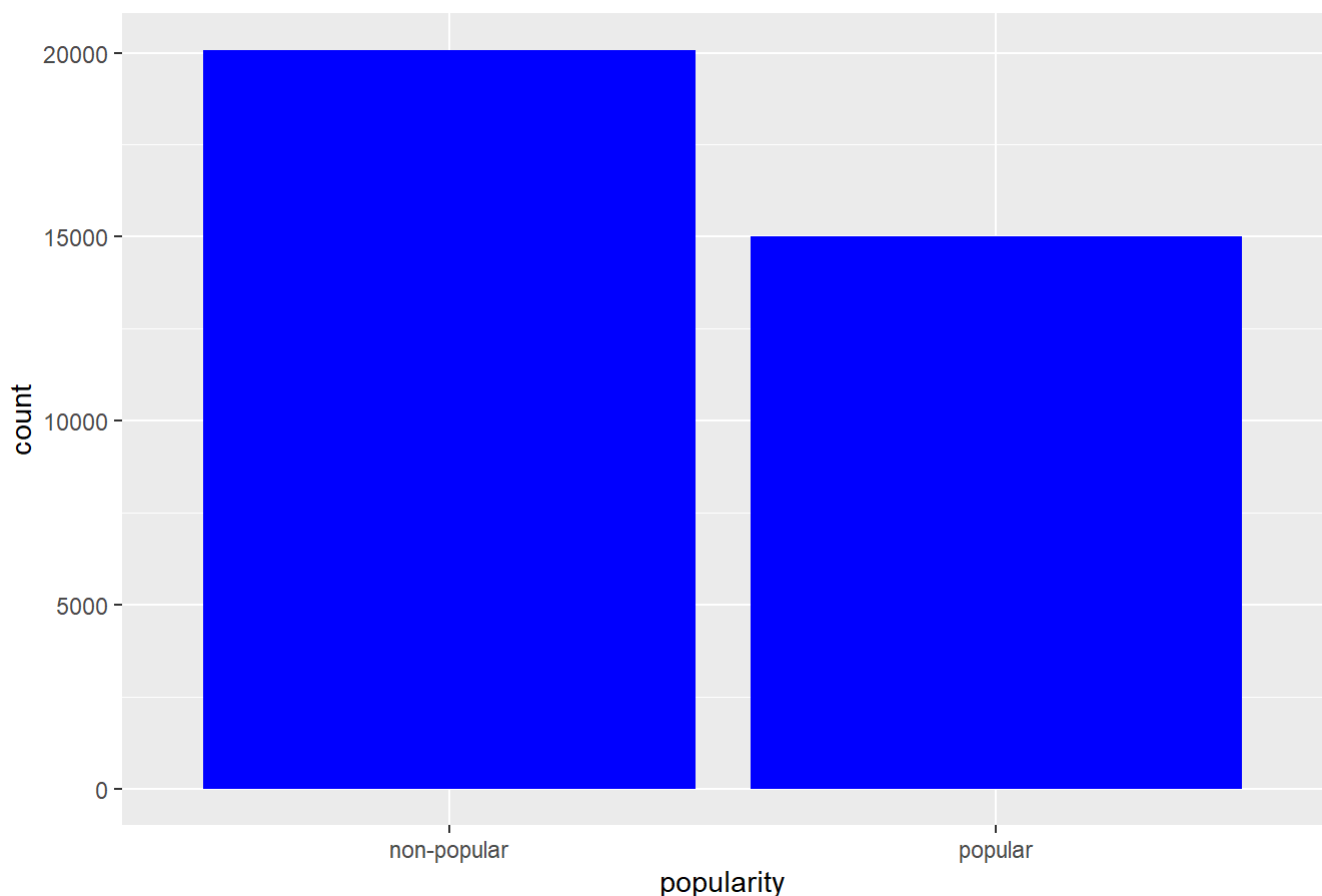
*# Displaying the frequency distribution of articles categorized into 'popular' and 'non-popular' classes*

```
library(ggplot2)
```

```
data_no_outliers$popularity <- ifelse(data_no_outliers$shares > 1400, "popular", "non-popular")
```

```
ggplot(data_no_outliers, aes(x = popularity)) + geom_bar(fill = "blue") + labs(title = "Frequency distribution of articles categorized into 'popular' and 'non-popular' classes")
```

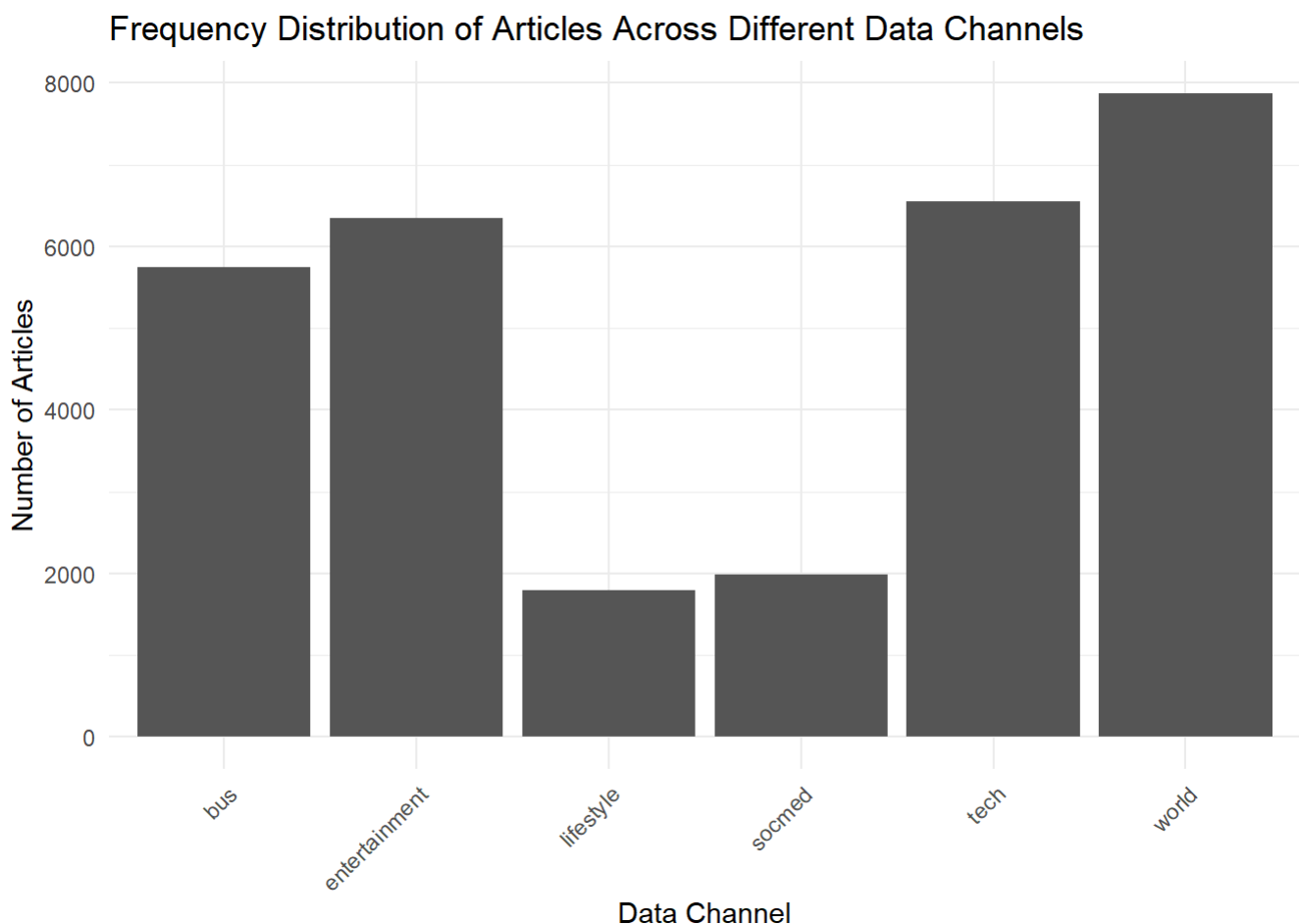
Frequency distribution of articles categorized into 'popular' and 'non-popular' classes



```
# Frequency distribution of articles across different data channels (Lifestyle, Entertainment, Business, Social Media, Technology, and World), distribute in onr graph
library(dplyr)
library(ggplot2)
library(tidyr)

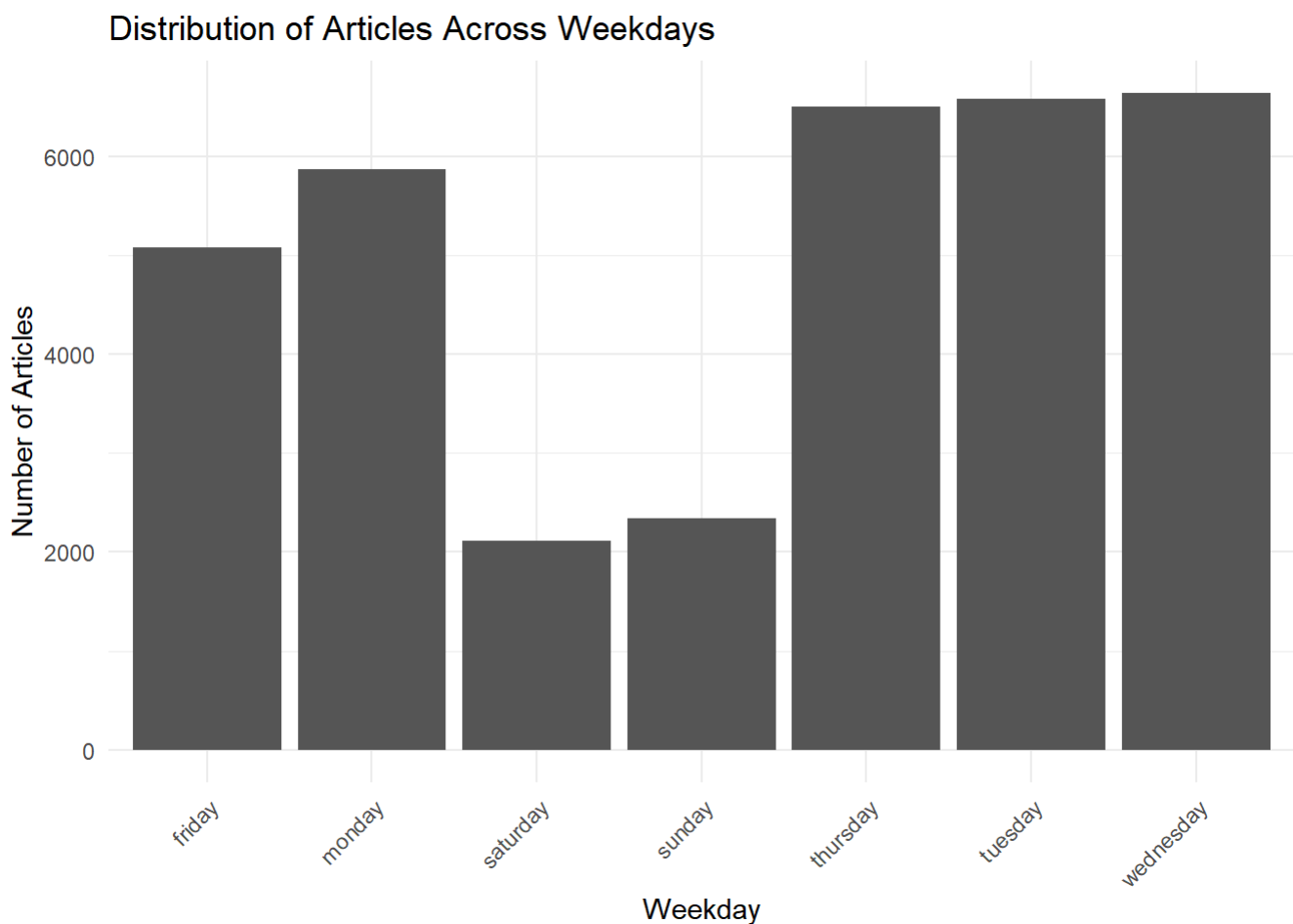
# Gather the data channel columns into key-value pairs
data_long <- data_no_outliers %>%
  select(starts_with("data_channel_is_")) %>%
  pivot_longer(cols = everything(), names_to = "Channel", values_to = "Frequency") %>%
  group_by(Channel) %>%
  summarise(Frequency = sum(Frequency)) %>%
  mutate(Channel = sub("data_channel_is_", "", Channel)) # Clean up the channel names

# Now, create the bar plot
ggplot(data_long, aes(x = Channel, y = Frequency)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  xlab("Data Channel") +
  ylab("Number of Articles") +
  ggtitle("Frequency Distribution of Articles Across Different Data Channels") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) # Rotate the x labels for better readability
```



```
# Frequency distribution of articles published on different weekdays (Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, and Sunday)
# Gather the weekday columns into key-value pairs
weekday_data_long <- data_no_outliers %>%
  select(starts_with("weekday_is_")) %>%
  pivot_longer(cols = everything(), names_to = "Weekday", values_to = "Frequency") %>%
  group_by(Weekday) %>%
  summarise(Frequency = sum(Frequency)) %>%
  mutate(Weekday = sub("weekday_is_", "", Weekday)) # Clean up the Weekday names

# Create the distribution graph for weekdays
ggplot(weekday_data_long, aes(x = Weekday, y = Frequency)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  xlab("Weekday") +
  ylab("Number of Articles") +
  ggtitle("Distribution of Articles Across Weekdays") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) # Rotate the x labels for readability
```



```

# Cross-tabulation of the number of articles published on different weekdays and channels
# Convert data channels from wide to long format
data_long <- data_no_outliers %>%
  pivot_longer(cols = starts_with("data_channel_is_"), names_to = "DataChannel", values_to =
"ChannelValue") %>%
  pivot_longer(cols = starts_with("weekday_is_"), names_to = "Weekday", values_to = "WeekdayV
alue") %>%
  filter(ChannelValue == 1, WeekdayValue == 1) %>%
  mutate(DataChannel = sub("data_channel_is_", "", DataChannel),
         Weekday = sub("weekday_is_", "", Weekday)) %>%
  select(-ChannelValue, -WeekdayValue)

# Ensure unique rows for the combination of articles, DataChannel, and Weekday
data_long <- distinct(data_long)

# Step 2: Cross-tabulation
cross_tab <- with(data_long, table(DataChannel, Weekday))

# View the cross-tabulation results
print(cross_tab)

```

```

##                Weekday
## DataChannel  friday monday saturday sunday thursday tuesday wednesday
##    bus              773   1049      207    306      1155   1083      1176
##  entertainment    875   1218      334   449      1113   1172      1186
##    lifestyle       267    275      152   173       306    289      336
##    socmed          276    282      162   111       400    398      360
##    tech            901   1107      453   332      1192   1311      1253
##    world          1234   1252      475   521      1467   1448      1481

```

### 3. Regression

First, we use random forest to predict the number of shares an article will get Random Forest is an ensemble learning method that is good for handling non-linear data with interactions between variables. It is robust to overfitting and can handle a large number of features. OOB error, which is an average error of predictions on each training sample  $x_i$ , using only the trees that did not have  $x_i$  in their bootstrap sample, can be used for model evaluation.

Advantages:

Good for complex datasets with non-linear relationships.

It has built-in methods for estimating model performance (OOB error).

It can handle missing data to a certain extent and doesn't require feature scaling.

Disadvantages:

Can be slow to train with very large datasets. Less interpretable than linear models.



```
# Random Forest Model
library(randomForest)

# Use parallel processing for faster training
library(doParallel)
cl <- makeCluster(detectCores() - 1) # Use all cores except one
registerDoParallel(cl)

data_no_outliers$url <- NULL # Remove the 'url' column
data_no_outliers$shares <- as.numeric(data_no_outliers$shares)

# Split the data into training and testing
set.seed(123) # Setting seed for reproducibility
train_indices <- sample(seq_len(nrow(data_no_outliers)), size = floor(0.8 * nrow(data_no_outliers)))
train_data <- data_no_outliers[train_indices, ]
test_data <- data_no_outliers[-train_indices, ]
train_data$popularity <- NULL
test_data$popularity <- NULL

rf_model <- randomForest(shares ~ ., data = train_data, ntree = 500)
# Print the model
rf_model
```

```
##
## Call:
## randomForest(formula = shares ~ ., data = train_data, ntree = 500)
##              Type of random forest: regression
##              Number of trees: 500
## No. of variables tried at each split: 19
##
##              Mean of squared residuals: 1048806
##              % Var explained: 13.4
```

```
# Prediction and RMSPE calculation
rf_predictions <- predict(rf_model, newdata = test_data)
rf_rmspe <- sqrt(mean((test_data$shares - rf_predictions)^2)) / mean(test_data$shares)
print(paste("RMSPE for Random Forest:", rf_rmspe))
```

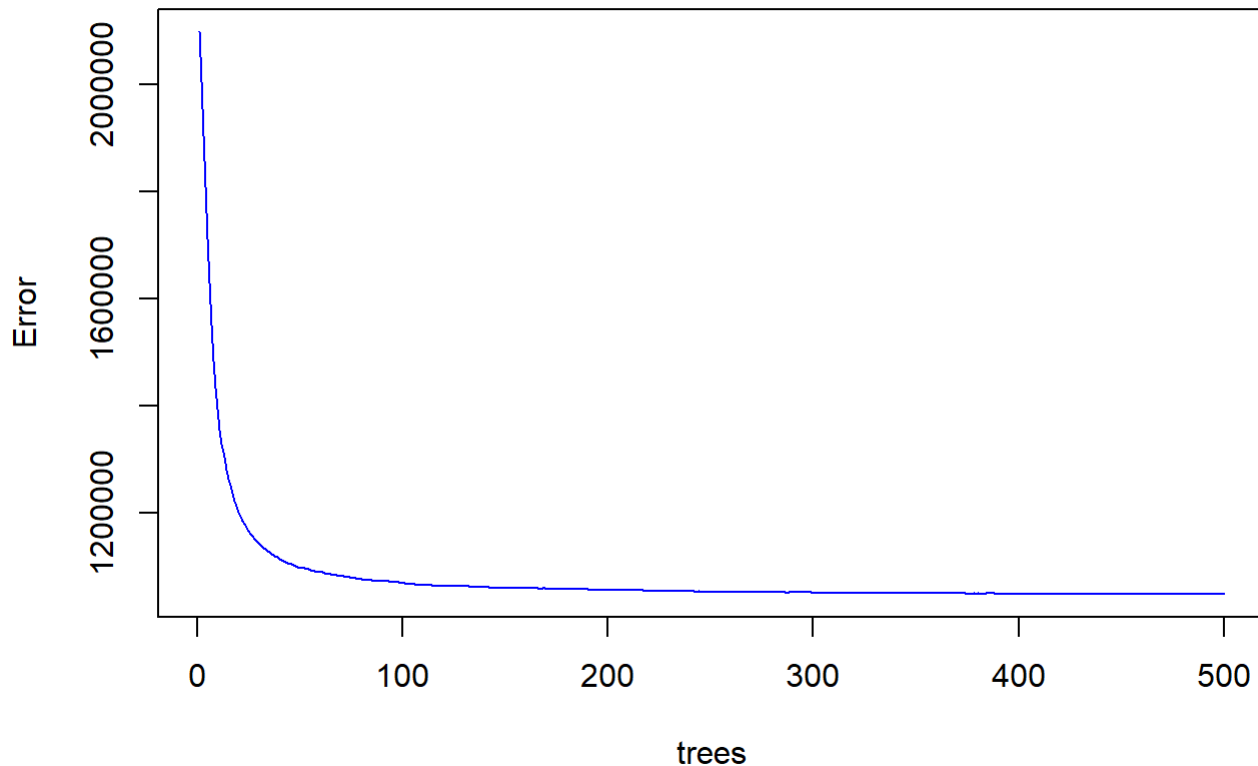
```
## [1] "RMSPE for Random Forest: 0.620274006828296"
```

```
# OOB
rf_oob <- rf_model$err.rate[nrow(rf_model$err.rate), "OOB"]
rf_oob
```

```
## NULL
```

```
plot(rf_model, col = "blue")
```

## rf\_model



# The second model we use is the GBM model.

GBM is a type of ensemble learning technique, where multiple models (typically decision trees) are combined to improve predictive performance. It builds the model in a stage-wise fashion; it constructs new models that predict the residuals or errors of prior models and then combines them into the final prediction.

Advantages:

GBM can deliver highly accurate predictions, often outperforming other machine learning algorithms.

It can be used for both classification and regression tasks.

GBM is capable of handling data with complex and non-linear relationships.

GBM naturally ranks features by their importance, which is useful for feature selection.

Disadvantages:

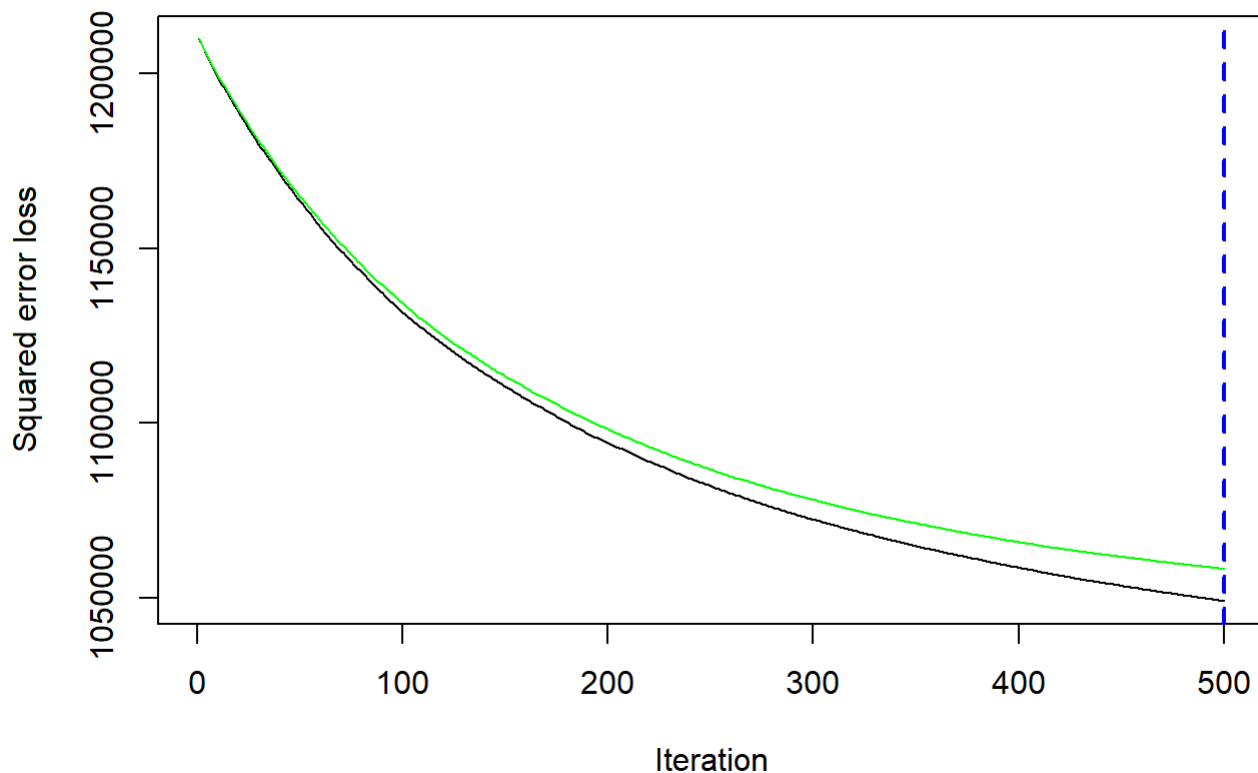
If not tuned properly, GBM models can easily overfit, especially on small datasets.

Training GBM models can be computationally intensive and time-consuming, especially with large datasets and a high number of trees.

There are several hyperparameters (like number of trees, depth of trees, learning rate) that need careful tuning to avoid overfitting and underfitting.

While individual trees are easy to interpret, a whole ensemble of trees (as used in GBM) can be difficult to interpret compared to simpler models.

```
# GBM Model
library(gbm)
set.seed(123) # Setting seed for reproducibility
# Exclude the 'popularity' variable from the training dataset
train_data$popularity <- NULL
# Fit the model - start with default settings and adjust as necessary
gbm_model <- gbm(shares ~ .,
  data = train_data,
  distribution = "gaussian", # for regression
  n.trees = 500,
  interaction.depth = 3,
  shrinkage = 0.01,
  cv.folds = 5, # for cross-validation
  n.minobsinnode = 10,
  verbose = FALSE)
best_trees <- gbm.perf(gbm_model, method = "cv")
```



```
# Predict on the test set
test_pred <- predict(gbm_model, newdata = test_data, n.trees = best_trees)
```

```
# OOB
gbm_oob <- gbm_model$train.error[best_trees]
gbm_oob
```

```
## [1] 1049020
```

```
# Calculate the RMSPE
gbm_rmspe <- sqrt(mean((test_data$shares - test_pred)^2)) / mean(test_data$shares)
print(paste("RMSPE for GBM:", gbm_rmspe))
```

```
## [1] "RMSPE for GBM: 0.625605120205234"
```

We use LPM as the benchmark model

```
# Linear Regression Model
lm_model <- lm(shares ~ ., data = train_data)
lm_predictions <- predict(lm_model, newdata = test_data)
lm_rmspe <- sqrt(mean((test_data$shares - lm_predictions)^2)) / mean(test_data$shares)

lm_rmspe
```

```
## [1] 0.6317187
```

Compare those models, we can see that The Random Forest model has the lowest RMSPE of approximately 62.02%, making it the best performer among the three in terms of prediction accuracy relative to the actual values.

```
# Improvements in predictive accuracy compared to the benchmark model
print(paste("Improvement in RMSPE over the benchmark model (Linear Regression):", rf_rmspe -
lm_rmspe))
```

```
## [1] "Improvement in RMSPE over the benchmark model (Linear Regression): -0.0114446739684028"
```

```
#Random Forest Feature Importance
# Extract feature importance
rf_importance <- importance(rf_model)

# Sort the importances in decreasing order, keeping the names attached
rf_sorted <- sort(rf_importance[, 1], decreasing = TRUE) # Make sure to select the correct i
mportance metric column

# Get the names of the top 10 most important features
top_features <- names(rf_sorted)[1:10]

# Print the names and importance scores of the top 10 features
print("Top 10 most important features in Random Forest:")
```

```
## [1] "Top 10 most important features in Random Forest:"
```

```
print(top_features)
```

```
## [1] "kw_avg_avg"          "kw_max_avg"
## [3] "timedelta"           "LDA_00"
## [5] "LDA_04"              "LDA_02"
## [7] "LDA_01"              "kw_avg_max"
## [9] "self_reference_min_shares" "LDA_03"
```

```
print(rf_sorted[1:10])
```

```
##          kw_avg_avg          kw_max_avg          timedelta
##          1536294821          1248049647          1106965489
##          LDA_00          LDA_04          LDA_02
##          1066104080          983860358          967967997
##          LDA_01          kw_avg_max self_reference_min_shares
##          957251273          945569798          924452507
##          LDA_03
##          898334227
```

*# Base on the feature importance above, we can create a bar plot to visualize the top 10 most important features*

*# Load the ggplot2 library*

```
library(ggplot2)
```

*# Create a vector of feature names*

```
features <- c("kw_avg_avg", "kw_max_avg", "timedelta", "LDA_02",
              "LDA_00", "LDA_04", "LDA_01", "self_reference_min_shares",
              "kw_avg_max", "LDA_03")
```

*# Create a vector of feature importance scores*

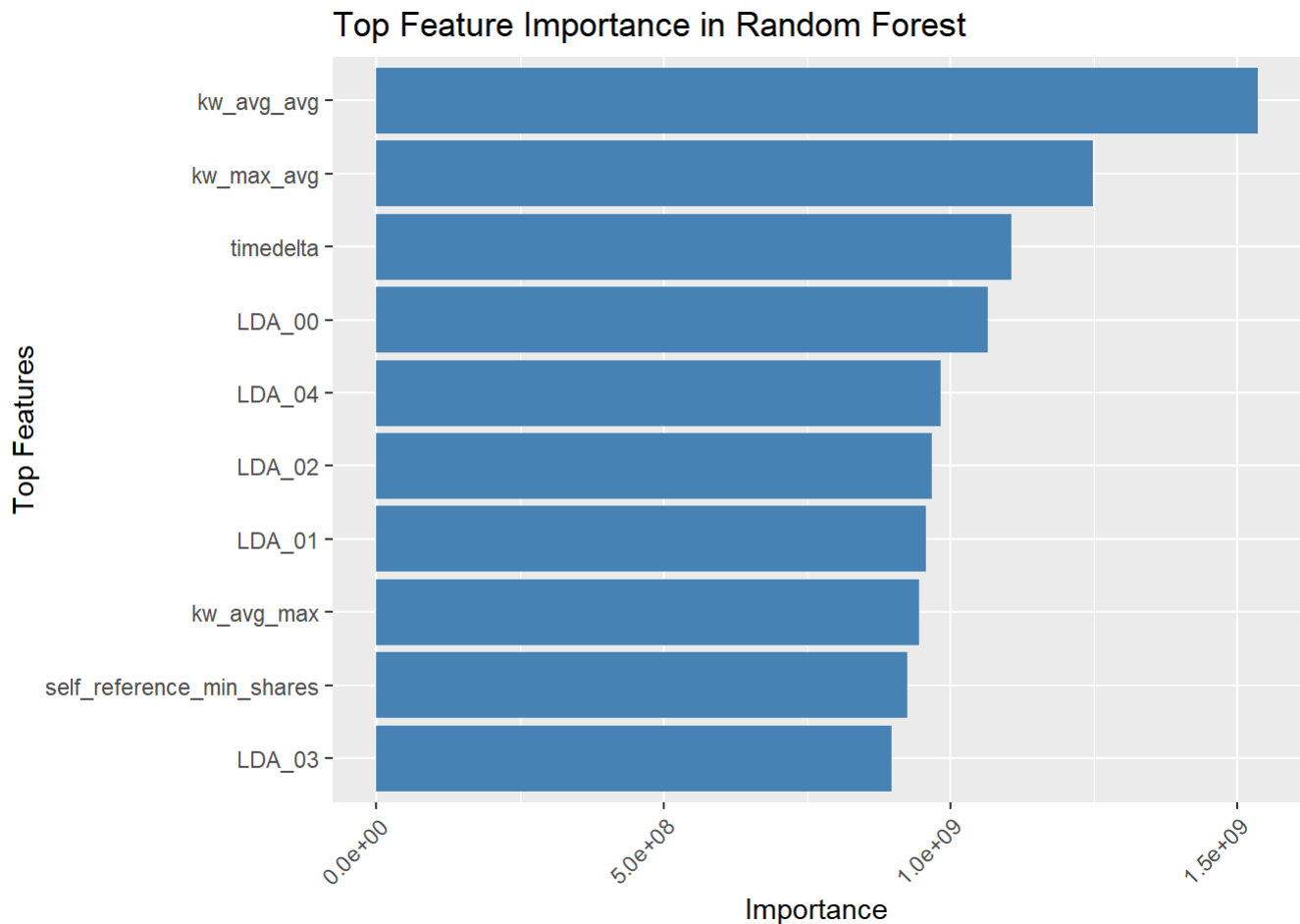
```
importances <- c(1536294821, 1248049647, 1106965489, 967967997,
                 1066104080, 983860358, 957251273, 924452507,
                 945569798, 898334227)
```

*# Combine into a data frame*

```
rf_imp_df <- data.frame(Feature = features, Importance = importances)
```

*# Create a bar plot for the top 10 most important features*

```
ggplot(rf_imp_df, aes(x = reorder(Feature, Importance), y = Importance)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "Top Feature Importance in Random Forest", x = "Top Features", y = "Importance") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1)) +
  coord_flip() # Flipping the coordinates for better readability
```



The Random Forest model outperformed the GBM and LPM models in terms of prediction accuracy, with an RMSPE of 62.02%. From this visualization, we can conclude that keyword-related features have a prominent role in the model's decisions, indicating the relevance of keyword metrics to the outcome variable. Additionally, several LDA (Latent Dirichlet Allocation) features also contribute meaningfully, suggesting that topics generated by LDA are useful for prediction in the given context. The feature timedelta also appears as an important factor, which may indicate that time-related aspects are significant in the model.

## 4. Classification

First, we use logistic regression to predict whether an article will be popular or not based on the number of shares it receives. Logistic regression is a statistical method used for binary classification, which predicts the probability that an observation belongs to one of two categories.

### Advantages:

Provides probabilities for outcomes, which are easy to interpret.

It is computationally inexpensive, making it highly efficient to train.

Offers good performance for linearly separable classes.

Gives a more informative result in terms of the likelihood of observations.

### Disadvantages:

It assumes a linear relationship between the logit of the outcome and each predictor variables.

It is typically restricted to binary classification problems.

Performance can be negatively impacted by unbalanced data, leading to biased models.

Susceptible to outliers and noisy data, which can distort the model's performance.

```
# Logistic Regression Model
# Convert the target variable 'shares' to a binary variable 'popularity'
data_no_outliers$popularity <- ifelse(data_no_outliers$shares > 1400, 1, 0)
data_no_outliers$shares <- NULL # Remove the 'shares' column

# Split the data into training and testing
set.seed(123) # Setting seed for reproducibility
train_indices <- sample(seq_len(nrow(data_no_outliers)), size = floor(0.8 * nrow(data_no_outliers)))
train_data <- data_no_outliers[train_indices, ]
test_data <- data_no_outliers[-train_indices, ]

# Fit the logistic regression model
logit_model <- glm(popularity ~ ., data = train_data, family = "binomial")
logit_model
```

```
##
## Call: glm(formula = popularity ~ ., family = "binomial", data = train_data)
##
## Coefficients:
##              (Intercept)              timedelta
##              -6.109e-01              -7.321e-05
##              n_tokens_title          n_tokens_content
##              -4.034e-03              1.717e-04
##              n_unique_tokens          n_non_stop_words
##              3.799e-02              -9.762e+00
##              n_non_stop_unique_tokens  num_hrefs
##              -8.553e-01              5.562e-03
##              num_self_hrefs          num_imgs
##              -1.966e-02              -2.591e-04
##              num_videos              average_token_length
##              -1.072e-03              -7.722e-02
##              num_keywords            data_channel_is_lifestyle
##              3.536e-02              -1.489e-01
## data_channel_is_entertainment        data_channel_is_bus
##              -2.836e-01              -2.257e-01
##              data_channel_is_socmed  data_channel_is_tech
##              8.272e-01              4.909e-01
##              data_channel_is_world   kw_min_min
##              -3.805e-02              1.627e-03
##              kw_max_min              kw_avg_min
##              4.468e-05              -2.979e-04
##              kw_min_max              kw_max_max
##              -4.343e-07              -4.309e-07
##              kw_avg_max              kw_min_avg
##              -5.889e-07              -5.288e-05
##              kw_max_avg              kw_avg_avg
##              -7.615e-05              5.870e-04
##              self_reference_min_shares self_reference_max_shares
##              2.490e-07              3.917e-07
##              self_reference_avg_sharess  weekday_is_monday
##              2.112e-06              -7.371e-01
##              weekday_is_tuesday         weekday_is_wednesday
##              -8.221e-01              -8.451e-01
##              weekday_is_thursday        weekday_is_friday
##              -7.720e-01              -5.785e-01
##              weekday_is_saturday        weekday_is_sunday
##              2.578e-01              NA
##              is_weekend                 LDA_00
##              NA                        6.412e-01
##              LDA_01                    LDA_02
##              -5.869e-01              -6.163e-01
##              LDA_03                    LDA_04
##              -5.622e-01              NA
##              global_subjectivity        global_sentiment_polarity
##              9.126e-01              -2.403e-01
##              global_rate_positive_words  global_rate_negative_words
##              -2.090e+00              2.916e+00
##              rate_positive_words         rate_negative_words
##              1.053e+01              1.002e+01
##              avg_positive_polarity       min_positive_polarity
```



```
##           -2.338e-01           -3.932e-01
##      max_positive_polarity      avg_negative_polarity
##           -2.200e-02           9.834e-02
##      min_negative_polarity      max_negative_polarity
##           3.184e-02           3.475e-02
##      title_subjectivity      title_sentiment_polarity
##           2.011e-01           2.332e-01
##      abs_title_subjectivity      abs_title_sentiment_polarity
##           2.957e-01           -1.034e-01
##
## Degrees of Freedom: 28081 Total (i.e. Null); 28025 Residual
## Null Deviance:      38340
## Residual Deviance: 35070      AIC: 35180
```

```
# Predictions and Evaluation
# Predict on the test set
logit_predictions <- predict(logit_model, newdata = test_data, type = "response")
```

```
## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
## prediction from rank-deficient fit; attr(*, "non-estim") has doubtful cases
```

```
logit_predictions <- ifelse(logit_predictions > 0.5, 1, 0)

# Confusion matrix
conf_matrix <- table(test_data$popularity, logit_predictions)
print(conf_matrix)
```

```
##      logit_predictions
##           0      1
##      0 3167  848
##      1 1553 1453
```

```
# Calculate accuracy
accuracy <- sum(diag(conf_matrix)) / sum(conf_matrix)
print(paste("Accuracy of Logistic Regression:", accuracy))
```

```
## [1] "Accuracy of Logistic Regression: 0.6580259222333"
```

```
# Calculate precision
precision <- conf_matrix[2, 2] / sum(conf_matrix[, 2])
print(paste("Precision of Logistic Regression:", precision))
```

```
## [1] "Precision of Logistic Regression: 0.631464580617123"
```

```
# Calculate recall
recall <- conf_matrix[2, 2] / sum(conf_matrix[2, ])
print(paste("Recall of Logistic Regression:", recall))
```

```
## [1] "Recall of Logistic Regression: 0.483366600133067"
```

```
# Calculate F1 score
f1_score <- 2 * (precision * recall) / (precision + recall)
print(paste("F1 Score of Logistic Regression:", f1_score))
```

```
## [1] "F1 Score of Logistic Regression: 0.547578669681553"
```

```
# Calculate AUC
library(pROC)
roc_obj <- roc(test_data$popularity, logit_predictions)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
auc <- auc(roc_obj)
print(paste("AUC of Logistic Regression:", auc))
```

```
## [1] "AUC of Logistic Regression: 0.636079315010494"
```

Then, we use decision tree model to do the same task. A decision tree consists of nodes that form a root, branches, and leaf nodes at the ends. Each internal node represents a “test” on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label (in classification) or a continuous value (in regression).

Advantages:

Trees can be visualized, making them easy to interpret even without statistical knowledge.

Can manage categorical and numerical data.

Doesn't assume a linear relationship and doesn't require much data preparation.

The features used at the top of the tree contribute the most to the decision process, indicating their importance.

Disadvantages:

Decision trees can create overly complex trees that do not generalize well to new data.

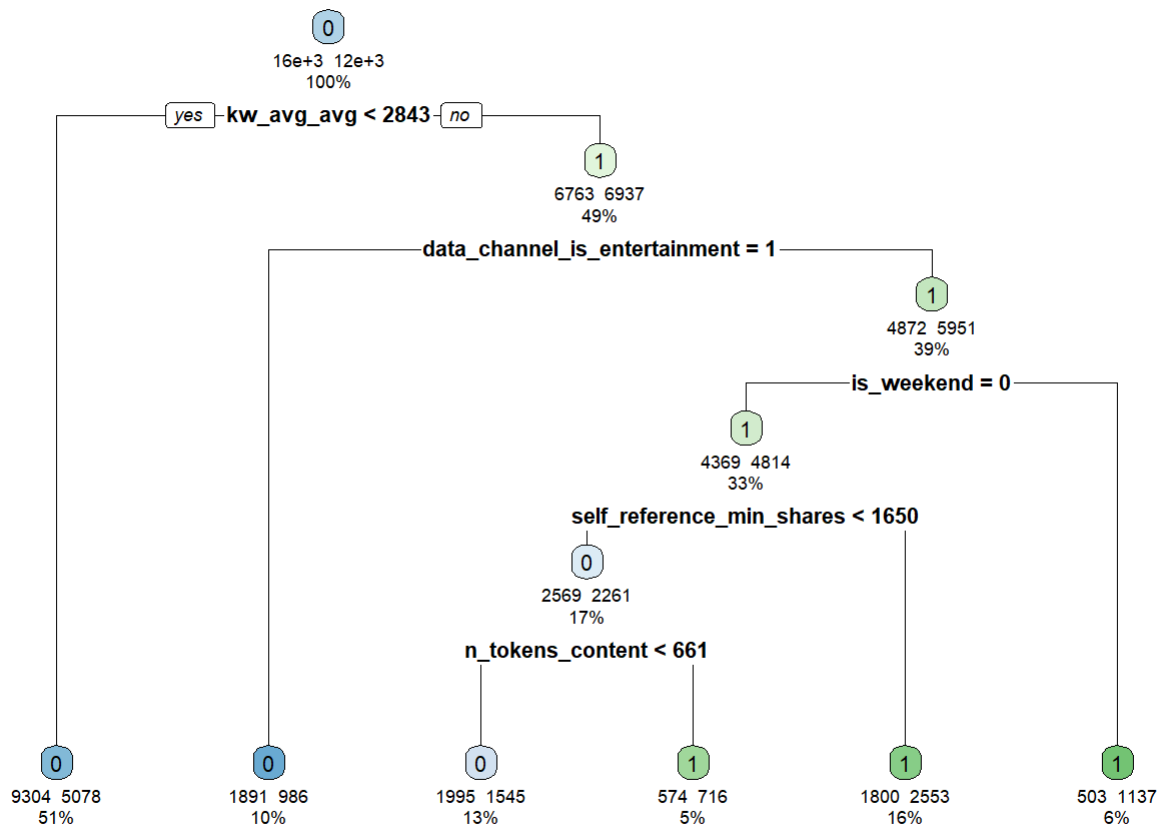
This can be mitigated by pruning.

Small changes in the data can result in a completely different tree being generated.

Trees can be biased toward classes that dominate if the class balance isn't addressed. They are not ideal for tasks where relationships between features are non-linear.

```
# Decision Tree Model
library(rpart)
library(rpart.plot)

# Fit the decision tree model
dt_model <- rpart(popularity ~ ., data = train_data, method = "class")
# Plot the decision tree
rpart.plot(dt_model, extra = 101, type = 2, fallen.leaves = TRUE, under = TRUE, facilen = 0)
```



```

# Predictions and Evaluation
# Predict on the test set
dt_predictions <- predict(dt_model, newdata = test_data, type = "class")

# Confusion matrix
conf_matrix_dt <- table(test_data$popularity, dt_predictions)
print(conf_matrix_dt)

```

```

##    dt_predictions
##      0      1
##  0 3312   703
##  1 1929 1077

```

```

# Calculate accuracy
accuracy_dt <- sum(diag(conf_matrix_dt)) / sum(conf_matrix_dt)
print(paste("Accuracy of Decision Tree:", accuracy_dt))

```

```
## [1] "Accuracy of Decision Tree: 0.625124626121635"
```

```

# Calculate precision
precision_dt <- conf_matrix_dt[2, 2] / sum(conf_matrix_dt[, 2])
print(paste("Precision of Decision Tree:", precision_dt))

```

```
## [1] "Precision of Decision Tree: 0.605056179775281"
```

```
# Calculate recall
recall_dt <- conf_matrix_dt[2, 2] / sum(conf_matrix_dt[2, ])
print(paste("Recall of Decision Tree:", recall_dt))
```

```
## [1] "Recall of Decision Tree: 0.358283433133733"
```

```
# Calculate F1 score
f1_score_dt <- 2 * (precision_dt * recall_dt) / (precision_dt + recall_dt)
print(paste("F1 Score of Decision Tree:", f1_score_dt))
```

```
## [1] "F1 Score of Decision Tree: 0.450062682824906"
```

```
# Calculate AUC
dt_predictions <- predict(dt_model, newdata = test_data, type = "prob")
positive_class_probabilities <- dt_predictions[, "1"]
# Calculate ROC and AUC
library(pROC)
roc_obj_dt <- roc(response = test_data$popularity, predictor = positive_class_probabilities)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
auc_value <- auc(roc_obj_dt)
print(auc_value)
```

```
## Area under the curve: 0.6119
```

Finally, Random forest model is used.

```
# Random Forest Model for classification
train_data$popularity <- as.factor(train_data$popularity)
# Fit the random forest model
rf_model <- randomForest(popularity ~ ., data = train_data, ntree = 500)
# Print the model
rf_model
```

```
##
## Call:
## randomForest(formula = popularity ~ ., data = train_data, ntree = 500)
##               Type of random forest: classification
##               Number of trees: 500
## No. of variables tried at each split: 7
##
## OOB estimate of error rate: 33.11%
## Confusion matrix:
##      0      1 class.error
## 0 12722 3345  0.2081907
## 1  5953 6062  0.4954640
```

```
# Predictions and Evaluation
# Predict on the test set
rf_predictions <- predict(rf_model, newdata = test_data)

# Confusion matrix
conf_matrix_rf <- table(test_data$popularity, rf_predictions)
print(conf_matrix_rf)
```

```
##      rf_predictions
##           0      1
##    0 3243   772
##    1 1462  1544
```

```
# Calculate accuracy
accuracy_rf <- sum(diag(conf_matrix_rf)) / sum(conf_matrix_rf)
print(paste("Accuracy of Random Forest:", accuracy_rf))
```

```
## [1] "Accuracy of Random Forest: 0.681811707733941"
```

```
# Calculate precision
precision_rf <- conf_matrix_rf[2, 2] / sum(conf_matrix_rf[, 2])
print(paste("Precision of Random Forest:", precision_rf))
```

```
## [1] "Precision of Random Forest: 0.666666666666667"
```

```
# Calculate recall
recall_rf <- conf_matrix_rf[2, 2] / sum(conf_matrix_rf[2, ])
print(paste("Recall of Random Forest:", recall_rf))
```

```
## [1] "Recall of Random Forest: 0.513639387890885"
```

```
# Calculate F1 score
f1_score_rf <- 2 * (precision_rf * recall_rf) / (precision_rf + recall_rf)
print(paste("F1 Score of Random Forest:", f1_score_rf))
```

```
## [1] "F1 Score of Random Forest: 0.580232995114619"
```

```
# Calculate AUC
rf_predictions <- predict(rf_model, newdata = test_data, type = "prob")
positive_class_probabilities_rf <- rf_predictions[, "1"]

# Calculate ROC and AUC
roc_obj_rf <- roc(response = test_data$popularity, predictor = positive_class_probabilities_rf)
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

```
auc_value_rf <- auc(roc_obj_rf)
print(auc_value_rf)
```

```
## Area under the curve: 0.7286
```

```
# Use LPM as the benchmark
lm_model <- lm(popularity ~ ., data = train_data)
```

```
## Warning in model.response(mf, "numeric"):
## 在因子响应上用type="numeric"的这一选项不会有效果
```

```
## Warning in Ops.factor(y, z$residuals): '-' not meaningful for factors
```

```
lm_predictions <- predict(lm_model, newdata = test_data)
lm_rmspe <- sqrt(mean((test_data$popularity - lm_predictions)^2)) / mean(test_data$popularity)
```

The Random Forest model has the highest values across all metrics, indicating that it is the best model among the ones compared here. It has the highest accuracy, precision, recall, F1 score, and AUC. The high AUC value of 0.729 suggests good discriminative ability of the Random Forest model.

```
# Improvements in predictive accuracy compared to the benchmark model
print(paste("Improvement in accuracy over the benchmark model (Logistic Regression):", accuracy_rf - accuracy))
```

```
## [1] "Improvement in accuracy over the benchmark model (Logistic Regression): 0.0237857855006409"
```

```
print(paste("Improvement in precision over the benchmark model (Logistic Regression):", precision_rf - precision))
```

```
## [1] "Improvement in precision over the benchmark model (Logistic Regression): 0.0352020860495437"
```

```
print(paste("Improvement in recall over the benchmark model (Logistic Regression):", recall_rf - recall))
```

```
## [1] "Improvement in recall over the benchmark model (Logistic Regression): 0.0302727877578177"
```

```
print(paste("Improvement in F1 score over the benchmark model (Logistic Regression):", f1_score_rf - f1_score))
```

```
## [1] "Improvement in F1 score over the benchmark model (Logistic Regression): 0.0326543254330658"
```

```
print(paste("Improvement in AUC over the benchmark model (Logistic Regression):", auc_value_rf - auc))
```

```
## [1] "Improvement in AUC over the benchmark model (Logistic Regression): 0.0925087558382612"
```

```
# Visualization of the confusion matrix for Random Forest  
library(caret)
```

```
## 载入需要的程辑包: lattice
```

```
confusionMatrix(conf_matrix_rf)
```

```
## Confusion Matrix and Statistics  
##  
##      rf_predictions  
##      0      1  
## 0 3243  772  
## 1 1462 1544  
##  
##              Accuracy : 0.6818  
##              95% CI : (0.6708, 0.6927)  
##      No Information Rate : 0.6701  
##      P-Value [Acc > NIR] : 0.01906  
##  
##              Kappa : 0.3309  
##  
##  McNemar's Test P-Value : < 2e-16  
##  
##              Sensitivity : 0.6893  
##              Specificity : 0.6667  
##              Pos Pred Value : 0.8077  
##              Neg Pred Value : 0.5136  
##              Prevalence : 0.6701  
##              Detection Rate : 0.4619  
##      Detection Prevalence : 0.5719  
##              Balanced Accuracy : 0.6780  
##  
##      'Positive' Class : 0  
##
```

```
# Feature Importance for Random Forest
# Extract feature importance
rf_importance <- importance(rf_model)
num_columns <- ncol(rf_importance)
# Sort the importances in decreasing order, keeping the names attached
if (num_columns >= 2) {
  # If there are at least 2 columns, sort the second one
  rf_sorted <- sort(rf_importance[, 2], decreasing = TRUE) # Assuming column 2 is the desired metric
} else if (num_columns == 1) {
  # If there's only one column, sort it
  rf_sorted <- sort(rf_importance[, 1], decreasing = TRUE) # Sorting the only column
} else {
  stop("The importance matrix does not have any columns.")
}
# Get the names of the top 10 most important features
top_features <- names(rf_sorted)[1:10]

# Print the names and importance scores of the top 10 features
print("Top 10 most important features in Random Forest:")
```

```
## [1] "Top 10 most important features in Random Forest:"
```

```
print(top_features)
```

```
## [1] "kw_avg_avg"          "kw_max_avg"
## [3] "timedelta"          "LDA_02"
## [5] "LDA_00"              "LDA_04"
## [7] "LDA_01"              "self_reference_min_shares"
## [9] "kw_avg_max"          "kw_avg_min"
```

```
print(rf_sorted[1:10])
```

```
##          kw_avg_avg          kw_max_avg          timedelta
##          505.1791          461.6429          419.6487
##          LDA_02          LDA_00          LDA_04
##          418.0953          408.6538          407.4231
##          LDA_01 self_reference_min_shares          kw_avg_max
##          394.6831          381.2182          374.1768
##          kw_avg_min
##          370.0082
```

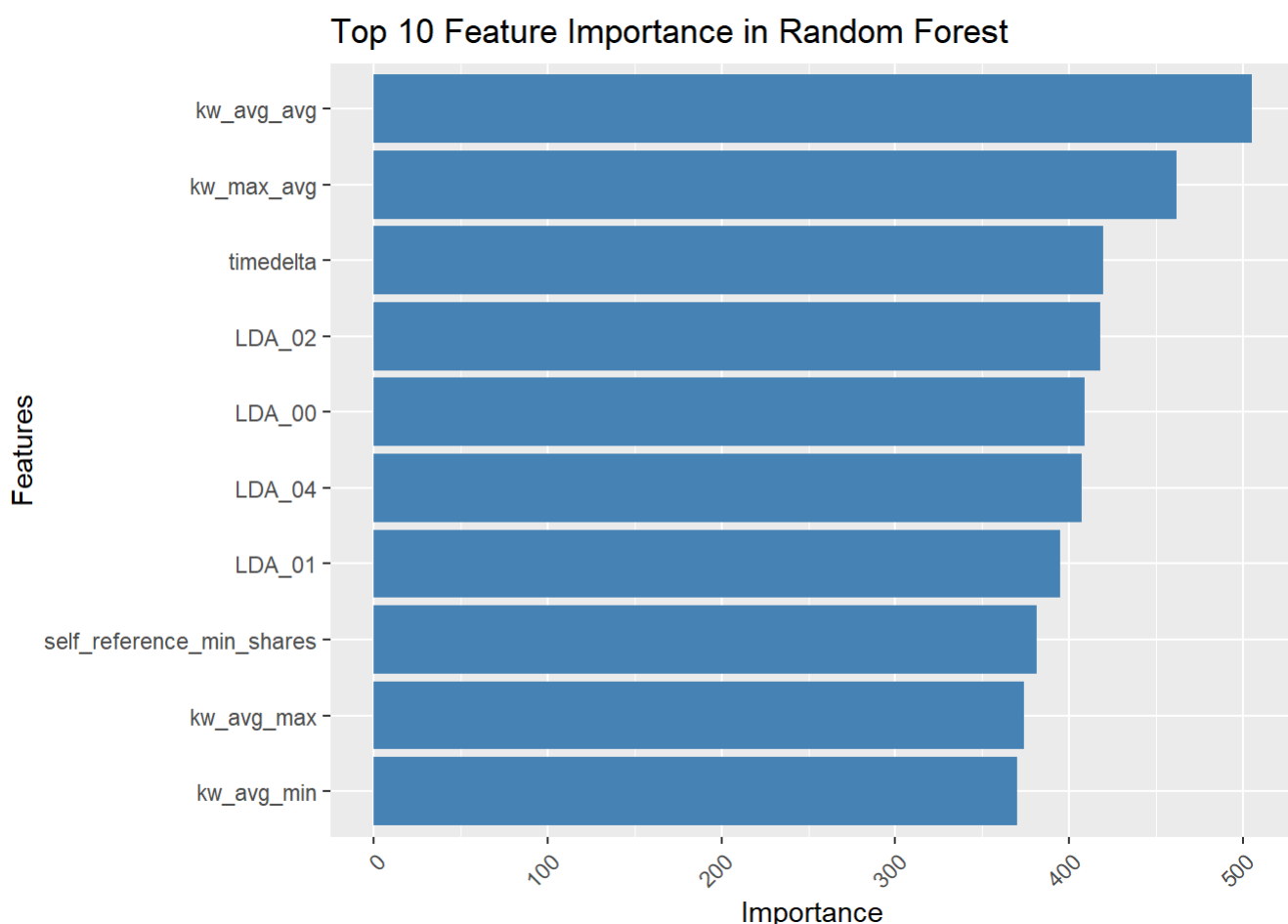


```
# Create a bar plot for the top 10 most important features
# Load the required library
library(ggplot2)

# Create a data frame for the top 10 features and their importance scores
features <- c("kw_avg_avg", "kw_max_avg", "timedelta", "LDA_02", "LDA_00",
             "LDA_04", "LDA_01", "self_reference_min_shares", "kw_avg_max", "kw_avg_min")
importances <- c(505.1791, 461.6429, 419.6487, 418.0953, 408.6538,
                407.4231, 394.6831, 381.2182, 374.1768, 370.0082)
feature_importance_df <- data.frame(Feature = features, Importance = importances)

# Order the data frame by Importance in descending order for plotting
feature_importance_df <- feature_importance_df[order(-feature_importance_df$Importance), ]

# Create a bar plot for the top 10 most important features
ggplot(feature_importance_df, aes(x = reorder(Feature, Importance), y = Importance)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "Top 10 Feature Importance in Random Forest", x = "Features", y = "Importance") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_flip() # Flip the coordinates for horizontal bar chart
```



From the chart, we can conclude the following:

**Keyword Average:** The feature `kw_avg_avg` stands out as the most influential in the model, indicating that the average keyword relevance has the strongest relationship with the target variable.

**Other Keyword Features:** `kw_max_avg`, `kw_avg_max`, and `kw_avg_min` also appear among the top important features, highlighting that keyword-related metrics play a significant role in the predictions made by the model.

**Timedelta:** The feature `timedelta` comes in as the third most important, suggesting that the time factor may have a substantial effect on the outcome.

**LDA Topics:** Several features generated from Latent Dirichlet Allocation (LDA), which are likely topic probabilities (`LDA_00`, `LDA_02`, `LDA_04`, `LDA_01`), are also crucial, demonstrating the model's reliance on the thematic structure of the data.

**Self-Reference:** The feature `self_reference_min_shares` is included in the list, indicating the model's sensitivity to how much the content references itself in terms of shares.

## 5. Recommendations

Based on the analysis and results obtained from the models, we can propose recommendations for content creators and marketers:

- Keyword Optimization:** Both models underscore the significance of keyword-related features (`kw_avg_avg`, `kw_max_avg`, `kw_avg_max`, `kw_avg_min`). Therefore, it's crucial to research and utilize keywords effectively within your content. Identify keywords that not only pertain to the subject matter but are also terms that your audience frequently searches for. These should be incorporated naturally into the title, headings, and throughout the content in a way that maintains readability and engagement.
- Topic Selection Using LDA:** The importance of LDA features (`LDA_00`, `LDA_01`, `LDA_02`, `LDA_04`) across both models indicates that the chosen topics of the articles significantly influence popularity. Employing topic modeling techniques like LDA can help identify prevalent themes within your domain that resonate with readers. Creating content that aligns with these topics or trends can increase relevance and engagement with your target audience.
- Timing of Content Release:** The `timedelta` feature's prominence suggests the timing of article publication can impact popularity. Analyze your audience's online behavior to identify when they are most active and likely to engage with new content. Schedule your posts accordingly to maximize visibility and shares. Consider different time zones and peak activity times based on your audience demographics.
- Self-Promotion and Social Proof:** `self_reference_min_shares` showing up in the classification model suggests that referencing one's own content and social share counts can positively affect popularity. Incorporate links to previous articles and display social share counts to provide social proof, potentially leading to increased trust and shareability.

## 6. Conclusion

In summary, the project's analysis of the Online News Popularity dataset aimed to model and predict article popularity. The Random Forest model exhibited the best performance with a notable AUC of 0.729, suggesting a strong ability to distinguish between popular and unpopular articles.

Feature importance highlighted that keywords and topics derived from LDA analysis, as well as the timing of publication ('`timedelta`'), are critical factors in predicting popularity. For content creators, this translates into a strategy that emphasizes keyword optimization, topical relevance, and timing of publication.

Challenges included managing the dataset's complexity and ensuring model generalization. Future research could explore more advanced text analysis techniques and the incorporation of real-time engagement data.

The project underscores the potential of machine learning to inform and enhance content creation strategies, ultimately aiding publishers in effectively engaging their audience.