[nltk\_data] Downloading package punkt to /root/nltk\_data...

[nltk\_data] Package punkt is already up-to-date!

[nltk\_data] Downloading package stopwords to /root/nltk\_data...

[nltk\_data] Package stopwords is already up-to-date!

[nltk\_data] Downloading package omw-1.4 to /root/nltk\_data...

[nltk\_data] Package omw-1.4 is already up-to-date!

DATAFRAME contains how many rows and columns:

(3306, 6)

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 3306 entries, 0 to 3305

Data columns (total 6 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Unnamed: 0 3306 non-null int64

1 Restaurant\_Name 3306 non-null object

2 Restaurant\_rated 3306 non-null float64

3 Review\_Written\_By 3306 non-null object

4 User\_Rated\_Restaurant 3306 non-null float64

5 Written\_Review 1909 non-null object

dtypes: float64(2), int64(1), object(3)

memory usage: 155.1+ KB

DATATYPES:

FIRST ROWS OF DATAFRAME ARE:

Unnamed: 0 Restaurant\_Name Restaurant\_rated Review\_Written\_By \

0 0 Toit 4.7 Brijmohan Patil

1 1 Toit 4.7 Jasleen Rajpal

2 2 Toit 4.7 Rukmini Das

3 3 Toit 4.7 Ritu Upadhyay

4 4 Toit 4.7 Sambhavi

User\_Rated\_Restaurant Written\_Review

0 4.0 NaN

1 5.0 Best food ever! And the packaging was a solid ...

2 4.0 crumb could be a litter lesser n fish more,non...

3 1.0 NaN

4 5.0 NaN

How many NULLS are their in zomato Dataframe:

Unnamed: 0 0

Restaurant\_Name 0

Restaurant\_rated 0

Review\_Written\_By 0

User\_Rated\_Restaurant 0

Written\_Review 1397

dtype: int64

Unnamed: 0 Restaurant\_Name Restaurant\_rated Review\_Written\_By \

0 0 Toit 4.7 Brijmohan Patil

1 1 Toit 4.7 Jasleen Rajpal

2 2 Toit 4.7 Rukmini Das

3 3 Toit 4.7 Ritu Upadhyay

4 4 Toit 4.7 Sambhavi

... ... ... ... ...

3301 3301 Cafe Paashh 4.9 Shuchi Saxena

3302 3302 Cafe Paashh 4.9 Mohnish Chordiya

3303 3303 Cafe Paashh 4.9 Aditya Shrimankar

3304 3304 Cafe Paashh 4.9 Nehaa T.

3305 3305 Cafe Paashh 4.9 Dishita Shah

User\_Rated\_Restaurant Written\_Review

0 4.0 -

1 5.0 Best food ever! And the packaging was a solid ...

2 4.0 crumb could be a litter lesser n fish more,non...

3 1.0 -

4 5.0 -

... ... ...

3301 5.0 -

3302 5.0 -

3303 3.0 -

3304 3.0 not too good. was expecting more from here sin...

3305 5.0 -

[3306 rows x 6 columns]

Splitted Review By Spaces:

0 1

1 11

2 16

3 1

4 1

..

3301 1

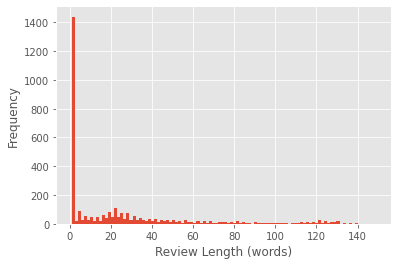
3302 1

3303 1

3304 13

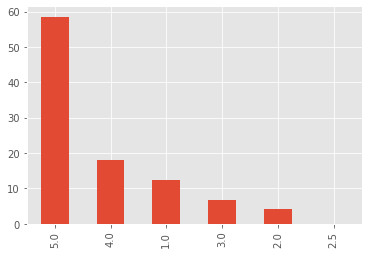
3305 1

Name: Written\_Review, Length: 3306, dtype: int64



Average words: 25.171203871748336

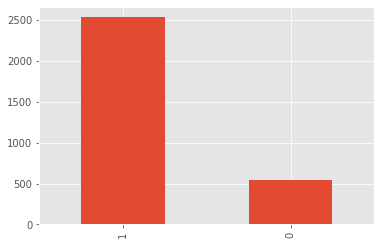
Skewness: 1.6621400523063772



1 2535

0 547

Name: Sentiment\_Rating, dtype: int64



After EDA dataframe has rows and columns: (3082, 7)

Number of unique tokens then: 7490

Number of unique tokens then: 6177

-Old Review:-

looks good, delivery boy spoiled it

-New Review-

looks good delivery boy spoiled it

Number of unique tokens then: 7490

Number of unique tokens then: 5512

['arabic', 'azerbaijani', 'basque', 'bengali', 'catalan', 'chinese', 'danish', 'dutch', 'english', 'finnish', 'french', 'german', 'greek', 'hebrew', 'hinglish', 'hungarian', 'indonesian', 'italian', 'kazakh', 'nepali', 'norwegian', 'portuguese', 'romanian', 'russian', 'slovene', 'spanish', 'swedish', 'tajik', 'turkish']

Original sentence best food ever and the packaging was a solid 10 10 loved it

Stopwords in the sentence ['and', 'the', 'was', 'a', 'it']

Non-stopwords in the sentence ['best', 'food', 'ever', 'packaging', 'solid', '10', '10', 'loved']

Written\_Review\_new \

0

1 best food ever and the packaging was a solid ...

2 crumb could be a litter lesser n fish more non...

3

4

... ...

3299 peaceful ambience variety of options in the m...

3300

3301

3302

3305

Written\_Review\_nonstop

0 []

1 [best, food, ever, packaging, solid, 10, 10, l...

2 [crumb, could, litter, lesser, n, fish, noneth...

3 []

4 []

... ...

3299 [peaceful, ambience, variety, options, menu, c...

3300 []

3301 []

3302 []

3305 []

[3082 rows x 2 columns]

[nltk\_data] Downloading package wordnet to /root/nltk\_data...

[nltk\_data] Package wordnet is already up-to-date!

Written\_Review\_nonstop \

0 []

1 [best, food, ever, packaging, solid, 10, 10, l...

2 [crumb, could, litter, lesser, n, fish, noneth...

3 []

4 []

... ...

3299 [peaceful, ambience, variety, options, menu, c...

3300 []

3301 []

3302 []

3305 []

Written\_Review\_stemm \

0 []

1 [best, food, ev, pack, solid, 10, 10, lov]

2 [crumb, could, lit, less, n, fish, nonetheless...

3 []

4 []

... ...

3299 [peac, amby, vary, opt, menu, coff, best, calm...

3300 []

3301 []

3302 []

3305 []

Written\_Review\_lemm

0 []

1 [best, food, ever, packaging, solid, 10, 10, l...

2 [crumb, could, litter, lesser, n, fish, noneth...

3 []

4 []

... ...

3299 [peaceful, ambience, variety, option, menu, co...

3300 []

3301 []

3302 []

3305 []

[3082 rows x 3 columns]

00 000 000rs 00pm 10 100 1000 100rs 1058 10pm ... yumm \

0 0 0 0 0 0 0 0 0 0 0 ... 0

1 0 0 0 0 2 0 0 0 0 0 ... 0

2 0 0 0 0 0 0 0 0 0 0 ... 0

3 0 0 0 0 0 0 0 0 0 0 ... 0

4 0 0 0 0 0 0 0 0 0 0 ... 0

yummiest yummmm yummy zafrani zero zest zishan zomato zone

0 0 0 0 0 0 0 0 0 0

1 0 0 0 0 0 0 0 0 0

2 0 0 0 0 0 0 0 0 0

3 0 0 0 0 0 0 0 0 0

4 0 0 0 0 0 0 0 0 0

[5 rows x 5481 columns]

(1, 601) 1

(1, 1962) 1

(1, 1735) 1

(1, 306) 1

(1, 4917) 1

(1, 3453) 1

(1, 5314) 1

(1, 123) 1

(1, 4505) 1

(1, 6) 2

(1, 2898) 1

(1, 2552) 1

(2, 306) 1

(2, 4917) 1

(2, 123) 1

(2, 1294) 1

(2, 1230) 1

(2, 552) 1

(2, 2841) 1

(2, 2788) 1

(2, 3197) 1

(2, 1923) 1

(2, 3135) 1

(2, 3274) 1

(2, 4909) 1

: :

(3077, 3638) 1

(3077, 5238) 1

(3077, 375) 1

(3077, 4864) 1

(3077, 4944) 1

(3077, 3322) 1

(3077, 2409) 1

(3077, 2452) 1

(3077, 283) 1

(3077, 3389) 1

(3077, 5267) 1

(3077, 1595) 1

(3077, 3946) 1

(3077, 5266) 1

(3077, 3189) 1

(3077, 434) 1

(3077, 3048) 1

(3077, 2555) 1

(3077, 3138) 1

(3077, 5212) 1

(3077, 2373) 1

(3077, 2267) 1

(3077, 832) 1

(3077, 3552) 1

(3077, 1071) 1

1 0.852512

0 0.147488

Name: Sentiment\_Rating, dtype: float64

MultinomialNB()

Prediction:

[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1

0 1 1 0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1

1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 0 1 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1

1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 0 1 1 0 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1

1 1 0 1 1 1 1 1 1 1 0 0 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 0 1 1

0 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1

1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1

1 1 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 1

1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1

1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]

Actual:

2623 1

2863 1

2152 1

2766 1

1136 1

..

1325 1

500 1

1965 1

470 1

1588 1

Name: Sentiment\_Rating, Length: 617, dtype: int64

CONFUSION MATRIX WITH BAG-OF-WORDS: [[ 54 9]

[ 37 517]]

OUR MODEL HAS ACCURACY WITH BAG-OF-WORDS: 92.54457050243113 %

F1 score: 0.9538745387453875

F1 score: 0.9505494505494506

words weights

61447 great 1.141768

59254 good 1.109278

9239 amazing 0.984232

48041 excellent 0.889265

22008 awesome 0.863657

92176 nice 0.828904

26621 best 0.752023

116117 rohit 0.663387

10355 ambience 0.657826

83428 love 0.471081

41128 delicious 0.463017

83661 loved 0.430832

133940 the best 0.406622

11813 and 0.396018

154223 vibe 0.390437

words weights

151250 unnecessarily expensive -0.590662

126767 stale food -0.608342

91817 never -0.615114

116464 rubbish -0.628857

116465 rubbish food -0.628857

153170 very bad -0.669963

67897 horrible -0.699630

126760 stale -0.736317

92903 no -0.821309

103919 pathetic -0.980776

36412 cold -1.029319

100196 order -1.044477

22594 bad -1.189304

93468 not -1.226441

167280 worst -2.042967

MultinomialNB()

Prediction:

[1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1

0 1 1 0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 0 1

1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 0 1 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1

1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1

1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 0 1 1 0 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1

1 1 0 1 1 1 1 1 1 1 0 0 1 0 1 1 1 1 0 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 0 1 1

0 0 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 1

1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1

1 1 1 0 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 0 1 1 1 1 1 1 1 1 1 0 1

1 0 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 0 1 1

1 1 1 0 1 1 1 1 0 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1]

Actual:

2623 1

2863 1

2152 1

2766 1

1136 1

..

1325 1

500 1

1965 1

470 1

1588 1

Name: Sentiment\_Rating, Length: 617, dtype: int64

CONFUSION MATRIX WITH TD-IDF: [[ 55 16]

[ 36 510]]

OUR MODEL HAS ACCURACY WITH TD-IDF: 91.57212317666126 %

F1 score TD-IDF: 0.9423423423423424