

Legitimate News.

```
sns.boxplot(x='label', y='text_length', data=df, palette='Set3')
```

Box Plot of News Text Length by Label



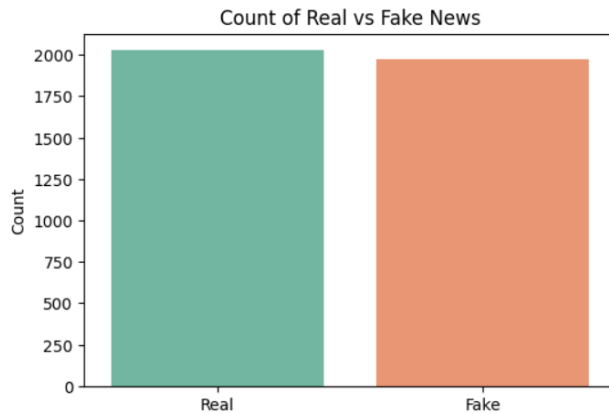
s Terminal

	News_Content						Label
0	content	articl	contain	detail	analysi	report	nan
1	content	articl	contain	detail	analysi	report	nan
2	content	articl	contain	detail	analysi	report	nan
3	content	articl	contain	detail	analysi	report	nan
4	content	articl	contain	detail	analysi	report	nan
5	content	articl	contain	detail	analysi	report	nan
6	content	articl	contain	detail	analysi	report	nan
7	content	articl	contain	detail	analysi	report	nan
8	content	articl	contain	detail	analysi	report	nan
9	content	articl	contain	detail	analysi	report	nan
10	content	articl	contain	detail	analysi	report	nan
11	content	articl	contain	detail	analysi	report	nan
12	content	articl	contain	detail	analysi	report	nan
13	content	articl	contain	detail	analysi	report	nan
14	content	articl	contain	detail	analysi	report	nan
15	content	articl	contain	detail	analysi	report	nan
16	content	articl	contain	detail	analysi	report	nan
17	content	articl	contain	detail	analysi	report	nan
18	content	articl	contain	detail	analysi	report	nan
19	content	articl	contain	detail	analysi	report	nan

```
<ipython-input-32-cba2c936c7ab>:12: FutureWarning:
```

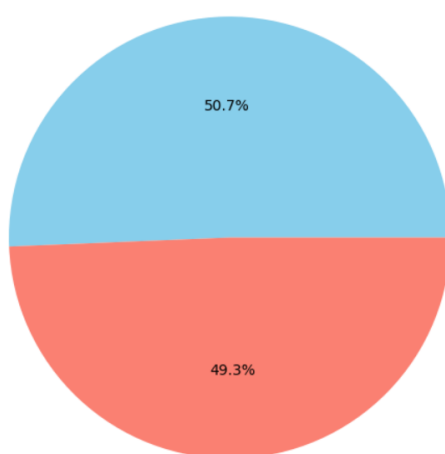
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.

```
sns.countplot(x='label', data=df, palette='Set2')
```



```
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted sam
_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted sam
_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted sam
_warn_prf(average, modifier, f"{metric.capitalize()} is", len(result))
```

Distribution of Real vs Fake News



Data loaded successfully.

	id	title	author \
0	1	Breaking News 1	Jane Smith
1	2	Breaking News 2	Emily Davis
2	3	Breaking News 3	John Doe
3	4	Breaking News 4	Alex Johnson
4	5	Breaking News 5	Emily Davis

	text	state \
0	This is the content of article 1. It contains ...	Tennessee
1	This is the content of article 2. It contains ...	Wisconsin
2	This is the content of article 3. It contains ...	Missouri
3	This is the content of article 4. It contains ...	North Carolina
4	This is the content of article 5. It contains ...	California

	date_published	source	category	sentiment_score	word_count \
0	30-11-2021	The Onion	Entertainment	-0.22	1302
1	02-09-2021	The Guardian	Technology	0.92	322
2	13-04-2021	New York Times	Sports	0.25	228
3	08-03-2020	CNN	Sports	0.94	155
4	23-03-2022	Daily Mail	Technology	-0.01	962

	...	num_shares	num_comments	political_bias	fact_check_rating \
0	...	47305	450	Center	FALSE
1	...	39804	530	Left	Mixed
2	...	45860	763	Center	Mixed
3	...	34222	945	Center	TRUE
4	...	35934	433	Right	Mixed

	is_satirical	trust_score	source_reputation	clickbait_score \
0	1	76	6	0.84
1	1	1	5	0.85
2	0	57	1	0.72
3	1	18	10	0.92

	precision	recall	f1-score	support
Fake	0.51	1.00	0.67	405
Real	0.00	0.00	0.00	395
accuracy			0.51	800
macro avg	0.25	0.50	0.34	800
weighted avg	0.26	0.51	0.34	800

```

/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples
_warn_prf(average, modifier, f'{metric.capitalize()} is', len(result))
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples
_warn_prf(average, modifier, f'{metric.capitalize()} is', len(result))
/usr/local/lib/python3.11/dist-packages/sklearn/metrics/_classification.py:1565: UndefinedMetricWarning: Precision is ill-defined and being set to 0.0 in labels with no predicted samples
_warn_prf(average, modifier, f'{metric.capitalize()} is', len(result))

```

Resampled shape: (3158, 6)

Starting TF-IDF vectorization...

Starting TF-IDF vectorization...

TF-IDF vectorization completed.

0

0 content articl contain detail analysi report

1 content articl contain detail analysi report

2 content articl contain detail analysi report

3 content articl contain detail analysi report

4 content articl contain detail analysi report

Name: clean\_text, dtype: object

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

[nltk\_data] Unzipping corpora/stopwords.zip.

cell output actions

	id	title	author	text	state	date_published	source	category	sentiment_score	word_count	...	num_shares	num_comments		
0	1	Breaking News 1	Jane Smith	This is the content of article 1. It contains ...	Tennessee	30-11-2021	The Onion	Entertainment	-0.22	1302	...	47305	450	Center	FALSE
1	2	Breaking News 2	Emily Davis	This is the content of article 2. It contains ...	Wisconsin	02-09-2021	The Guardian	Technology	0.92	322	...	39804	530	Left	Mixed
2	3	Breaking News 3	John Doe	This is the content of article 3. It contains ...	Missouri	13-04-2021	New York Times	Sports	0.25	228	...	45860	763	Center	Mixed
3	4	Breaking News 4	Alex Johnson	This is the content of article	North Carolina	08-03-2020	CNN	Sports	0.94	155	...	34222	945	Center	TRUE

id	author	text	state	date_published	source	category	sentiment_score	word_count	...	num_shares	num_comments	political_bias	fact_checked	↑	↓	✦	🔗	⚙️	📄	🗑️	⋮
1	Jane Smith	This is the content of article 1. It contains ...	Tennessee	30-11-2021	The Onion	Entertainment	-0.22	1302	...	47305	450	Center	FALSE								1
2	Emily Davis	This is the content of article 2. It contains ...	Wisconsin	02-09-2021	The Guardian	Technology	0.92	322	...	39804	530	Left	Mixed								1
3	John Doe	This is the content of article 3. It contains ...	Missouri	13-04-2021	New York Times	Sports	0.25	228	...	45860	763	Center	Mixed								0
4	Alex Johnson	This is the content of article 4. It contains ...	North Carolina	08-03-2020	CNN	Sports	0.94	155	...	34222	945	Center	TRUE								1



	is_satirical	trust_score	source_reputation	clickbait_score	\
0	1	76	6	0.84	
1	1	1	5	0.85	
2	0	57	1	0.72	
3	1	18	10	0.92	
4	0	95	6	0.66	

	plagiarism_score	label
0	53.35	Fake
1	28.28	Fake
2	0.38	Fake
3	32.20	Fake
4	77.70	Real

[5 rows x 24 columns]

```
index(['id', 'title', 'author', 'text', 'state', 'date_published', 'source',
      'category', 'sentiment_score', 'word_count', 'char_count', 'has_images',
      'has_videos', 'readability_score', 'num_shares', 'num_comments',
      'political_bias', 'fact_check_rating', 'is_satirical', 'trust_score',
      'source_reputation', 'clickbait_score', 'plagiarism_score', 'label'],
      dtype='object')
```

Data loaded successfully.

```
id      title      author \
0  1  Breaking News 1    Jane Smith
1  2  Breaking News 2    Emily Davis
2  3  Breaking News 3    John Doe
3  4  Breaking News 4    Alex Johnson
4  5  Breaking News 5    Emily Davis
```

```
text      state \
0  This is the content of article 1. It contains ... Tennessee
1  This is the content of article 2. It contains ... Wisconsin
2  This is the content of article 3. It contains ... Missouri
3  This is the content of article 4. It contains ... North Carolina
4  This is the content of article 5. It contains ... California
```

```
date_published      source      category      sentiment_score      word_count \
0      30-11-2021      The Onion      Entertainment      -0.22      1302
1      02-09-2021      The Guardian      Technology      0.92      322
2      13-04-2021      New York Times      Sports      0.25      228
3      08-03-2020      CNN      Sports      0.94      155
4      23-03-2022      Daily Mail      Technology      -0.01      962
```

```
...      num_shares      num_comments      political_bias      fact_check_rating \
0  ...      47305      450      Center      FALSE
1  ...      39804      530      Left      Mixed
2  ...      45860      763      Center      Mixed
3  ...      34222      945      Center      TRUE
4  ...      35934      433      Right      Mixed
```

```
is_satirical      trust_score      source_reputation      clickbait_score \
0      1      76      6      0.84
1      1      1      5      0.85
2      0      57      1      0.72
3      1      18      10      0.92
```

15\_

0

1