# Fake_News_Analysis_EDA

October 5, 2025

### 0.0.1 Exploratory Data Analysis on Fake News Dataset

[2]:

```
import pandas as pd
```

[3]:

```
Buzzfeed_f = pd.read_csv("data/BuzzFeed_fake_news_content.csv")
Buzzfeed_r = pd.read_csv("data/BuzzFeed_real_news_content.csv")

gossipcop_f =  pd.read_csv("data/gossipcop_fake.csv")
gossipcop_r =  pd.read_csv("data/gossipcop_real.csv")

politifact_f =  pd.read_csv("data/politifact_fake.csv")
politifact_r =  pd.read_csv("data/politifact_real.csv")
```

[6]:

```
Buzzfeed_f.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 91 entries, 0 to 90
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   id              91 non-null     object
 1   title           91 non-null     object
 2   text            91 non-null     object
 3   url             83 non-null     object
 4   top_img         81 non-null     object
 5   authors         57 non-null     object
 6   source          83 non-null     object
 7   publish_date    77 non-null     object
 8   movies          21 non-null     object
 9   images          81 non-null     object
 10  canonical_link  80 non-null     object
 11  meta_data       91 non-null     object
dtypes: object(12)
memory usage: 8.7+ KB
```

[11]:

```
gossipcop_f.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5323 entries, 0 to 5322
```

```
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   id         5323 non-null   object
 1   news_url   5067 non-null   object
 2   title      5323 non-null   object
 3   tweet_ids  5135 non-null   object
dtypes: object(4)
memory usage: 166.5+ KB
```

```
[12]: politifact_f.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 432 entries, 0 to 431
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   id         432 non-null    object
 1   news_url   428 non-null    object
 2   title      432 non-null    object
 3   tweet_ids  392 non-null    object
dtypes: object(4)
memory usage: 13.6+ KB
```

```
[4]: Buzzfeed_merge=pd.concat([Buzzfeed_r,Buzzfeed_f],axis=0)

     Buzzfeed_merge['news_type']=Buzzfeed_merge['id'].apply(lambda x: x.
      ↪split('_')[0])
     Buzzfeed_merge.head()
```

```
[4]:                 id                                             title  \
     0   Real_1-Webpage  Another Terrorist Attack in NYC…Why Are we STI…
     1  Real_10-Webpage  Donald Trump: Drugs a 'Very, Very Big Factor' …
     2  Real_11-Webpage  Obama To UN: ‘Giving Up Liberty, Enhances Secu…
     3  Real_12-Webpage  Trump vs. Clinton: A Fundamental Clash over Ho…
     4  Real_13-Webpage  President Obama Vetoes 9/11 Victims Bill, Sett…

                                                     text  \
     0  On Saturday, September 17 at 8:30 pm EST, an e…
     1  Less than a day after protests over the police…
     2  Obama To UN: ‘Giving Up Liberty, Enhances Secu…
     3  Getty Images Wealth Of Nations Trump vs. Clint…
     4  President Obama today vetoed a bill that would…

                                                      url  \
     0  http://eaglerising.com/36942/another-terrorist…
     1                              http://abcn.ws/2d41Nn9
     2  http://rightwingnews.com/barack-obama/obama-un…
```

```
3                          http://politi.co/2de2qs0
4                            http://abcn.ws/2dh2NFs

                                             top_img  \
0  http://eaglerising.com/wp-content/uploads/2016…
1  http://a.abcnews.com/images/Politics/AP_donald…
2  http://rightwingnews.com/wp-content/uploads/20…
3  http://static.politico.com/e9/11/6144cdc24e319…
4  http://a.abcnews.com/images/US/AP_Obama_BM_201…

                                            authors                    source  \
0                   View All Posts,Leonora Cravotta    http://eaglerising.com
1       More Candace,Adam Kelsey,Abc News,More Adam            http://abcn.ws
2                                       Cassy Fiano  http://rightwingnews.com
3        Jack Shafer,Erick Trickey,Zachary Karabell          http://politi.co
4  John Parkinson,More John,Abc News,More Alexander            http://abcn.ws

               publish_date                                     movies  \
0  {'$date': 1474528230000}                                        NaN
1                       NaN                                        NaN
2  {'$date': 1474476044000}  https://www.youtube.com/embed/ji6pl5Vwrvk
3  {'$date': 1474974420000}                                        NaN
4                       NaN                                        NaN

                                              images  \
0  http://constitution.com/wp-content/uploads/201…
1  http://www.googleadservices.com/pagead/convers…
2  http://rightwingnews.com/wp-content/uploads/20…
3  https://static.politico.com/dims4/default/8a1c…
4  http://www.googleadservices.com/pagead/convers…

                                      canonical_link  \
0  http://eaglerising.com/36942/another-terrorist…
1  http://abcnews.go.com/Politics/donald-trump-dr…
2  http://rightwingnews.com/barack-obama/obama-un…
3  http://www.politico.com/magazine/story/2016/09…
4  http://abcnews.go.com/Politics/president-obama…

                                           meta_data news_type
0  {"description": "\u201cWe believe at this poin…      Real
1  {"fb_title": "Trump: Drugs a 'Very, Very Big F…      Real
2  {"googlebot": "noimageindex", "og": {"site_nam…      Real
3  {"description": "He sees it as zero-sum. She b…      Real
4  {"fb_title": "President Obama Vetoes 9/11 Vict…      Real
```

```
[6]: Buzzfeed_merge.shape
```

[6]: (182, 13)

[7]:
```
Buzzfeed_merge.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 182 entries, 0 to 90
Data columns (total 13 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   id              182 non-null    object
 1   title           182 non-null    object
 2   text            182 non-null    object
 3   url             174 non-null    object
 4   top_img         172 non-null    object
 5   authors         141 non-null    object
 6   source          174 non-null    object
 7   publish_date    133 non-null    object
 8   movies          25 non-null     object
 9   images          172 non-null    object
 10  canonical_link  170 non-null    object
 11  meta_data       182 non-null    object
 12  news_type       182 non-null    object
dtypes: object(13)
memory usage: 19.9+ KB
```

[9]:
```
Buzzfeed_merge['contain_movies']=Buzzfeed_merge['movies'].apply(lambda x: 0 if␣
 ↪str(x)=='nan' else 1)
Buzzfeed_merge['contain_images']=Buzzfeed_merge['images'].apply(lambda x: 0 if␣
 ↪str(x)=='nan' else 1)

Buzzfeed_drop = Buzzfeed_merge.drop(['id','url',
                     'top_img',
                     'authors',
                     'publish_date',
                     'canonical_link',
                     'meta_data',
                     'movies',
                     'images'],axis=1)

Buzzfeed_drop.head()
```

[9]:
```
                                               title  \
0  Another Terrorist Attack in NYC…Why Are we STI…
1  Donald Trump: Drugs a 'Very, Very Big Factor' …
2  Obama To UN: ‘Giving Up Liberty, Enhances Secu…
3  Trump vs. Clinton: A Fundamental Clash over Ho…
4  President Obama Vetoes 9/11 Victims Bill, Sett…
```

```
                                              text  \
0  On Saturday, September 17 at 8:30 pm EST, an e…
1  Less than a day after protests over the police…
2  Obama To UN: ‘Giving Up Liberty, Enhances Secu…
3  Getty Images Wealth Of Nations Trump vs. Clint…
4  President Obama today vetoed a bill that would…

                     source news_type  contain_movies  contain_images
0    http://eaglerising.com      Real               0               1
1            http://abcn.ws      Real               0               1
2  http://rightwingnews.com      Real               1               1
3          http://politi.co      Real               0               1
4            http://abcn.ws      Real               0               1
```

[10]:

```python
Buzzfeed_drop.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 182 entries, 0 to 90
Data columns (total 6 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   title           182 non-null    object
 1   text            182 non-null    object
 2   source          174 non-null    object
 3   news_type       182 non-null    object
 4   contain_movies  182 non-null    int64
 5   contain_images  182 non-null    int64
dtypes: int64(2), object(4)
memory usage: 10.0+ KB
```

[11]:

```python
Buzzfeed_clean = Buzzfeed_drop["source"].fillna(0)
```

[14]:

```python
# Save the DataFrame to a CSV file
output_directory = 'data/'  # Replace with your actual path
filename = 'Buzzfeed_data.csv'
full_path = output_directory + filename

Buzzfeed_clean.to_csv(full_path, index=False)
```