

Github URL: <https://github.com/Drukyi54/PythonCode>

##Import Libraries

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

##Ingest data from csv

```
df=pd.read_csv("/content/data_news.csv")
```

```
df.head()
```

	Unnamed: 0		title	desc
\				
0	0	The Inside Story of Microsoft's Partnership wi...		NaN
1	1	Microsoft secures nonvoting board seat at OpenAI		NaN
2	2	Microsoft Needs a Better Seat at OpenAI's Table		NaN
3	3	AWS unveils 'Amazon Q' AI assistant, jabs at M...		NaN
4	4	AI wars: Amazon is coming for Microsoft and Go...		NaN

	date	datetime	\
0	3 days ago	2023-12-01 07:16:16.417645	
1	5 days ago	2023-11-29 07:16:16.418681	
2	Nov 27	NaN	
3	6 days ago	2023-11-28 07:16:16.420499	
4	4 days ago	2023-11-30 07:16:16.421417	

	link	\
0	news.google.com/articles/CBMiZGh0dHBz0i8vd3d3L...	
1	news.google.com/articles/CBMiV2h0dHBz0i8vd3d3L...	
2	news.google.com/articles/CBMiU2h0dHBz0i8vd3d3L...	
3	news.google.com/articles/CBMiaWh0dHBz0i8vd3d3L...	
4	news.google.com/articles/CBMiX2h0dHBz0i8vZmluY...	

	img	
media	\	
0	https://lh3.googleusercontent.com/proxy/uS28Fi...	The New Yorker
1	https://lh3.googleusercontent.com/proxy/mp0zLA...	CNBC
2	https://lh3.googleusercontent.com/proxy/sz4Lch...	The Wall Street Journal
3	https://lh3.googleusercontent.com/proxy/Fh0tzA...	

GeekWire

4 <https://lh3.googleusercontent.com/proxy/AklVbo...>

Yahoo

Finance

	site
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

##Manage different datatypes and data wrangling

Drop unnamed files

```
df.drop(columns=['Unnamed: 0'], inplace=True)
```

```
df.info()
```

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

```
RangeIndex: 100 entries, 0 to 99
```

```
Data columns (total 8 columns):
```

#	Column	Non-Null Count	Dtype
0	title	100 non-null	object
1	desc	0 non-null	float64
2	date	100 non-null	object
3	datetime	85 non-null	object
4	link	100 non-null	object
5	img	88 non-null	object
6	media	100 non-null	object
7	site	0 non-null	float64

```
dtypes: float64(2), object(6)
```

```
memory usage: 6.4+ KB
```

```
df.describe()
```

	desc	site
count	0.0	0.0
mean	NaN	NaN
std	NaN	NaN
min	NaN	NaN
25%	NaN	NaN
50%	NaN	NaN
75%	NaN	NaN
max	NaN	NaN

Drop the NaN column names "desc" and "site"

```
columns = ['desc', 'site']
df.drop(columns=columns, inplace=True)

df.head()
```

	title	date \
0	The Inside Story of Microsoft's Partnership wi...	3 days ago
1	Microsoft secures nonvoting board seat at OpenAI	5 days ago
2	Microsoft Needs a Better Seat at OpenAI's Table	Nov 27
3	AWS unveils 'Amazon Q' AI assistant, jabs at M...	6 days ago
4	AI wars: Amazon is coming for Microsoft and Go...	4 days ago

	datetime \
0	2023-12-01 07:16:16.417645
1	2023-11-29 07:16:16.418681
2	NaN
3	2023-11-28 07:16:16.420499
4	2023-11-30 07:16:16.421417

	link \
0	news.google.com/articles/CBMiZGh0dHBz0i8vd3d3L...
1	news.google.com/articles/CBMiV2h0dHBz0i8vd3d3L...
2	news.google.com/articles/CBMiU2h0dHBz0i8vd3d3L...
3	news.google.com/articles/CBMiaWh0dHBz0i8vd3d3L...
4	news.google.com/articles/CBMiX2h0dHBz0i8vZmluY...

	img	media
0	https://lh3.googleusercontent.com/proxy/uS28Fi...	The New Yorker
1	https://lh3.googleusercontent.com/proxy/mp0zLA...	CNBC
2	https://lh3.googleusercontent.com/proxy/sz4Lch...	The Wall Street Journal
3	https://lh3.googleusercontent.com/proxy/Fh0tzA...	GeekWire
4	https://lh3.googleusercontent.com/proxy/AklVbo...	Yahoo Finance

```
print(df.isnull().sum())
```

```
title      0
date       0
datetime   15
link       0
img        12
media      0
dtype: int64
```

```
df_cleaned = df.dropna()
df_cleaned
```

	title	date \
0	The Inside Story of Microsoft's Partnership wi...	3 days ago
1	Microsoft secures nonvoting board seat at OpenAI	5 days ago
3	AWS unveils 'Amazon Q' AI assistant, jabs at M...	6 days ago
4	AI wars: Amazon is coming for Microsoft and Go...	4 days ago
5	Amazon Q: AWS' answer to Microsoft's GPT-drive...	6 days ago
..
92	Direct Travel Unveils New Integration Between ...	4 days ago
93	Microsoft and IMDA to upskill 2,000 SMEs in Si...	6 days ago
94	Increases to Microsoft Corporation's (NASDAQ:M...	3 days ago
95	Sam Altman-OpenAI board crisis: Microsoft CEO ...	2 days ago
97	RapidFort's Platform Now Available in Microsof...	6 days ago
	datetime \	
0	2023-12-01 07:16:16.417645	
1	2023-11-29 07:16:16.418681	
3	2023-11-28 07:16:16.420499	
4	2023-11-30 07:16:16.421417	
5	2023-11-28 07:16:16.422292	
..	...	
92	2023-11-30 07:16:16.501674	
93	2023-11-28 07:16:16.502547	
94	2023-12-01 07:16:16.503424	
95	2023-12-02 07:16:16.504321	
97	2023-11-28 07:16:16.506050	
	link \	
0	news.google.com/articles/CBMiZGh0dHBz0i8vd3d3L...	
1	news.google.com/articles/CBMiV2h0dHBz0i8vd3d3L...	
3	news.google.com/articles/CBMiaWh0dHBz0i8vd3d3L...	
4	news.google.com/articles/CBMiX2h0dHBz0i8vZmluY...	
5	news.google.com/articles/CBMiaGh0dHBz0i8vd3d3L...	
..	...	
92	news.google.com/articles/CBMiV2h0dHBz0i8vZmluY...	
93	news.google.com/articles/CBMihwFodHRwczovL25ld...	
94	news.google.com/articles/CBMiWmh0dHBz0i8vZmluY...	
95	news.google.com/articles/CBMiuQFodHRwczovL3d3d...	
97	news.google.com/articles/CBMiWGh0dHBz0i8vZmluY...	
	img	media
0	https://lh3.googleusercontent.com/proxy/uS28Fi...	The New Yorker
1	https://lh3.googleusercontent.com/proxy/mp0zLA...	CNBC
3	https://lh3.googleusercontent.com/proxy/Fh0tzA...	GeekWire
4	https://lh3.googleusercontent.com/proxy/AklVbo...	Yahoo Finance
5	https://lh3.googleusercontent.com/proxy/1UXow3...	Computerworld

```

..
...
92 https://lh3.googleusercontent.com/proxy/Ec285w... Yahoo Finance
93 https://lh3.googleusercontent.com/proxy/HZX_li... Microsoft
94 https://lh3.googleusercontent.com/proxy/yUAjFX... Yahoo Finance
95 https://lh3.googleusercontent.com/proxy/3bkkz1... Business Today
97 https://lh3.googleusercontent.com/proxy/3BAjCF... Yahoo Finance

[74 rows x 6 columns]

```

##visualize data

Below is the visualization of news published by different media

```

plt.figure(figsize=(18, 8))
sns.countplot(data=df_cleaned, x='media')
plt.title('Distribution of Articles by Media')
plt.xlabel('Media')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()

```



