

```
!pip install pandas
import pandas as pd
import numpy as np
from functools import reduce

Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (1.5.3)
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas) (2023.4)
Requirement already satisfied: numpy>=1.21.0 in /usr/local/lib/python3.10/dist-packages (from pandas) (1.25.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.1->pandas) (1.16.0)
```

```
df=pd.read_csv('/content/BL-Flickr-Images-Book.csv')
df.head()
```

| | Identifier | Edition Statement | Place of Publication | Date of Publication | Publisher | Title | Author | Contributors | Corporate Author | Corporate Contributors | Former owner | Er |
|---|------------|------------------------------|--------------------------|---------------------|-----------------------|---|-----------|--|------------------|------------------------|--------------|----|
| 0 | 206 | NaN | London | 1879 [1878] | S. Tinsley & Co. | Walter Forbes. [A novel.] By A. A. | A. A. | FORBES, Walter. | NaN | NaN | NaN | |
| 1 | 216 | NaN | London; Virtue & Yorston | 1868 | Virtue & Co. | All for Greed. [A novel. The dedication signed... | A., A. A. | BLAZE DE BURY, Marie Pauline Rose - Baroness | NaN | NaN | NaN | |
| 2 | 218 | NaN | London | 1869 | Bradbury, Evans & Co. | Love the Avenger. By the author of "All for Gr... | A., A. A. | BLAZE DE BURY, Marie Pauline Rose - Baroness | NaN | NaN | NaN | |
| 3 | 472 | NaN | London | 1851 | James Darling | Welsh Sketches, chiefly ecclesiastical, to the... | A., E. S. | Appleyard, Ernest Silvanus. | NaN | NaN | NaN | |
| 4 | 480 | A new edition, revised, etc. | London | 1857 | Wertheim & Macintosh | [The World in which I live, and my place in it... | A., E. S. | BROOME, John Henry. | NaN | NaN | NaN | |

```
to_drop=['Edition Statement' , 'Contributors', 'Corporate Author', 'Corporate Contributors']
df.drop(to_drop, inplace = True,axis=1)
```

```
df.head()
```

| | Identifier | Place of Publication | Date of Publication | Publisher | Title | Author |
|---|------------|--------------------------|---------------------|-----------------------|---|-----------|
| 0 | 206 | London | 1879 [1878] | S. Tinsley & Co. | Walter Forbes. [A novel.] By A. A. | A. A. |
| 1 | 216 | London; Virtue & Yorston | 1868 | Virtue & Co. | All for Greed. [A novel. The dedication signed... | A., A. A. |
| 2 | 218 | London | 1869 | Bradbury, Evans & Co. | Love the Avenger. By the author of "All for Gr... | A., A. A. |
| 3 | 472 | London | 1851 | James Darling | Welsh Sketches, chiefly ecclesiastical, to the... | A., E. S. |
| 4 | 480 | London | 1857 | Wertheim & Macintosh | [The World in which I live, and my place in it... | A., E. S. |

```
df.set_index('Identifier',inplace=True)
df.head ()
```

| | Place of Publication | Date of Publication | Publisher | Title | Author |
|------------|--------------------------|---------------------|-----------------------|---|-----------|
| Identifier | | | | | |
| 206 | London | 1879 [1878] | S. Tinsley & Co. | Walter Forbes. [A novel.] By A. A. | A. A. |
| 216 | London; Virtue & Yorston | 1868 | Virtue & Co. | All for Greed. [A novel. The dedication signed... | A., A. A. |
| 218 | London | 1869 | Bradbury, Evans & Co. | Love the Avenger. By the author of "All for Gr... | A., A. A. |
| 472 | London | 1851 | James Darling | Welsh Sketches, chiefly ecclesiastical, to the... | A., E. S. |
| 480 | London | 1857 | Wertheim & Macintosh | [The World in which I live, and my place in it... | A., E. S. |

```
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 8287 entries, 206 to 4160339
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Place of Publication    8287 non-null   object
1   Date of Publication     8106 non-null   object
2   Publisher               4092 non-null   object
3   Title                  8287 non-null   object
4   Author                 6509 non-null   object
dtypes: object(5)
memory usage: 388.5+ KB

```

```
df['Date of Publication'].head(25)
```

```

Identifier
206      1879 [1878]
216      1868
218      1869
472      1851
480      1857
481      1875
519      1872
667      NaN
874      1676
1143     1679
1280     1802
1808     1859
1905     1888
1929     1839, 38-54
2836     1897
2854     1865
2956     1860-63
2957     1873
3017     1866
3131     1899
4598     1814
4884     1820
4976     1800
5382     1847, 48 [1846-48]
5385     [1897?]
Name: Date of Publication, dtype: object

```

```

u_cha=['[',',','-']
def clean_dates(item):
    dop=str(item.loc['Date of Publication'])

    if dop == 'nan' or dop[0] == '[':
        return np.NaN

    for cha in u_cha:
        if cha in dop:
            cha_index=dop.find(cha)
            dop=dop[:cha_index]
    return dop
df['Date of Publication']=df.apply(clean_dates,axis=1)

```

```

def clean_aun(author):
    author =str(author)
    if author == 'nan':
        return 'NaN'
    author = author.split(',')

    if len(author) ==1 :
        name=filter(lambda x:x.isalpha(),author[0])
        return reduce(lambda x,y:x+y,name)

    l_name,f_name=author[0],author[1]
    f_name=f_name[:f_name.find('-')] if '-'in f_name else f_name
def clean_t(title):
    if title == 'nan':
        return 'NaN'
    if title[0]=='[':
        title=title[1:title.find('[')]

    if 'by' in title:
        title=title[1:title.find('by')]

    elif 'By' in title:
        title=title[1:title.find('By')]

    if ']'in title:
        title=title[1:title.find(')]')]

    title=title[:-2]
    title =list(map(str.capitalize,title.split()))
    return ' '.join(title)
df['Title']=df['Title'].apply(clean_t)
df.head()

```

| Identifier | Place of Publication | Date of Publication | Publisher | Title | Author |
|------------|--------------------------|---------------------|-----------------------|---|--------|
| 206 | London | 1879 | S. Tinsley & Co. | Lter Forbes. [a Nove | AA |
| 216 | London; Virtue & Yorston | 1868 | Virtue & Co. | LI For Greed. [a Novel. The Dedication Signed:... | A. AA. |
| 218 | London | 1869 | Bradbury, Evans & Co. | Ove The Avenger | A. AA. |
| 472 | London | 1851 | James Darling | Elsh Sketches, Chiefly Ecclesiastical, To The ... | E. SA. |
| 480 | London | 1857 | Wertheim & Macintosh | | E. SA. |