ex333aa

```python
import pandas as pd
import numpy as np

# Import the data into a DataFrame
df = pd.read_csv('BL-Flickr-Images-Book.csv')

# Display the first few rows of the DataFrame
print("Original DataFrame:")
print(df.head())

# Find and drop the columns which are irrelevant for the
book information
irrelevant_columns = ['Edition Statement', 'Corporate Author'
, 'Corporate Contributors', 'Former owner', 'Engraver',
'Contributors', 'Issuance type', 'Shelfmarks']
df.drop(columns=irrelevant_columns, inplace=True)

# Change the Index of the DataFrame
df.set_index('Identifier', inplace=True)

# Tidy up fields in the data such as date of publication with
the help of simple regular expression
df['Date of Publication'] = df['Date of Publication'].str.extract(r
'^(\d{4})', expand=False)

# Combine str methods with NumPy to clean columns
df['Place of Publication'] = np.where(df['Place of Publication'
].str.contains('London'), 'London', df['Place of Publication'].
str.replace('-', ' '))

# Display the cleaned DataFrame
print("\nCleaned DataFrame:")
print(df.head())
```

Read More