09_handling_bad_data

March 21, 2022

1 Handling Bad, Missing, and Duplicate Data

Abid Ali
https://github.com/abid-2362

1.0.1 Stripping whitespaces

```python
import pandas as pd

data = pd.read_csv('./data/artwork_data.csv', low_memory=False)
data.head()
```

```
[ ]:      id accession_number          artist artistRole  artistId  \
    0  1035          A00001    Blake, Robert     artist        38
    1  1036          A00002    Blake, Robert     artist        38
    2  1037          A00003    Blake, Robert     artist        38
    3  1038          A00004    Blake, Robert     artist        38
    4  1039          A00005  Blake, William     artist        39

                                                  title               dateText  \
    0  A Figure Bowing before a Seated Old Man with h…          date not known
    1  Two Drawings of Frightened Figures, Probably f…          date not known
    2  The Preaching of Warning. Verso: An Old Man En…                  ?c.1785
    3    Six Drawings of Figures with Outstretched Arms          date not known
    4  The Circle of the Lustful: Francesca da Rimini…  1826-7, reprinted 1892

                                         medium  \
    0  Watercolour, ink, chalk and graphite on paper…
    1                              Graphite on paper
    2      Graphite on paper. Verso: graphite on paper
    3                              Graphite on paper
    4                      Line engraving on paper

                                       creditLine  year  acquisitionYear  \
    0          Presented by Mrs John Richmond 1922   NaN           1922.0
    1          Presented by Mrs John Richmond 1922   NaN           1922.0
    2          Presented by Mrs John Richmond 1922  1785           1922.0
    3          Presented by Mrs John Richmond 1922   NaN           1922.0
```

1

```
4  Purchased with the assistance of a special gra…  1826         1919.0
```

```
               dimensions width height  depth units inscription  \
0  support: 394 x 419 mm    394    419    NaN    mm         NaN
1  support: 311 x 213 mm    311    213    NaN    mm         NaN
2  support: 343 x 467 mm    343    467    NaN    mm         NaN
3  support: 318 x 394 mm    318    394    NaN    mm         NaN
4    image: 243 x 335 mm    243    335    NaN    mm         NaN
```

```
   thumbnailCopyright                                   thumbnailUrl  \
0                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
1                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
2                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
3                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
4                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
```

```
                                             url
0  http://www.tate.org.uk/art/artworks/blake-a-fi…
1  http://www.tate.org.uk/art/artworks/blake-two-…
2  http://www.tate.org.uk/art/artworks/blake-the-…
3  http://www.tate.org.uk/art/artworks/blake-six-…
4  http://www.tate.org.uk/art/artworks/blake-the-…
```

```python
[ ]: data.loc[data.title.str.contains('\s$', regex=True)]
```

```
[ ]:            id accession_number                    artist artistRole  artistId  \
     49498    4308           P07466  Hamilton Finlay, Ian      artist      1093
     50534    2235           P11065         Clarke, Brian      artist       911
     50535    2236           P11066         Clarke, Brian      artist       911
     50537    2238           P11068         Clarke, Brian      artist       911
     53186   21168           P77679     Bourgeois, Louise      artist      2351
     56283    5826           T00705     Hamilton, Richard      artist      1244
     67409   88191           T12064          Leach, David      artist      7651
     67432   88215           T12087        Leach, Bernard      artist      1478
```

```
                                                  title dateText  \
     49498  Port-Distinguishing Letters of Scottish Fishin…    1976
     50534                                             Boys    1981
     50535                                        Buildings    1981
     50537                                   Pray for Josquin    1981
     53186                            Untitled (Safety Pins)    1991
     56283  Towards a definitive statement on the coming t…    1962
     67409                          2 Standard Ware Mead Cups  1945-55
     67432                                             Bowl   c.1960
```

```
                                     medium  \
     49498            Screenprint on ceramic tile
```

```
                                                   medium
50534                                   Screenprint on paper
50535                                   Screenprint on paper
50537                                   Screenprint on paper
53186                                     Drypoint on paper
56283   Oil paint, cellulose paint and printed paper o…
67409                                         Ochre porcelain
67432                                               Porcelain

                                           creditLine  year  \
49498                                Purchased 1981  1976
50534                    Presented by Paul Beldock 1983  1981
50535                    Presented by Paul Beldock 1983  1981
50537                    Presented by Paul Beldock 1983  1981
53186                                Purchased 1994  1991
56283                                Purchased 1964  1962
67409   Accepted by HM Government in lieu of inheritan…  1945
67432   Accepted by HM Government in lieu of inheritan…  1960

       acquisitionYear                                dimensions  \
49498           1981.0                    unconfirmed: 153 x 153 mm
50534           1983.0                       image: 1003 x 700 mm
50535           1983.0                       image: 1005 x 698 mm
50537           1983.0                       image: 1115 x 688 mm
53186           1994.0                        image: 303 x 379 mm
56283           1964.0  support: 610 x 813 mm frame: 809 x 1011 x 810 mm
67409           2005.0                     object: 70 x 80 x 80 mm
67432           2005.0                    object: 50 x 140 x 140 mm

       width  height  depth units     inscription  \
49498    153     153    NaN    mm             NaN
50534   1003     700    NaN    mm  date inscribed
50535   1005     698    NaN    mm  date inscribed
50537   1115     688    NaN    mm  date inscribed
53186    303     379    NaN    mm             NaN
56283    610     813    NaN    mm  date inscribed
67409     70      80   80.0    mm             NaN
67432     50     140  140.0    mm             NaN

                                    thumbnailCopyright  \
49498                 © Estate of Ian Hamilton Finlay
50534   © Brian Clarke. All Rights Reserved 2014 / DACS
50535   © Brian Clarke. All Rights Reserved 2014 / DACS
50537   © Brian Clarke. All Rights Reserved 2014 / DACS
53186                          © The Easton Foundation
56283             © The estate of Richard Hamilton
67409               © The estate of Bernard Leach
67432               © The estate of Bernard Leach
```

```
                                                       thumbnailUrl  \
49498  http://www.tate.org.uk/art/images/work/P/P07/P…
50534  http://www.tate.org.uk/art/images/work/P/P11/P…
50535  http://www.tate.org.uk/art/images/work/P/P11/P…
50537  http://www.tate.org.uk/art/images/work/P/P11/P…
53186  http://www.tate.org.uk/art/images/work/P/P77/P…
56283  http://www.tate.org.uk/art/images/work/T/T00/T…
67409  http://www.tate.org.uk/art/images/work/T/T12/T…
67432  http://www.tate.org.uk/art/images/work/T/T12/T…


                                                       url
49498  http://www.tate.org.uk/art/artworks/hamilton-f…
50534  http://www.tate.org.uk/art/artworks/clarke-boy…
50535  http://www.tate.org.uk/art/artworks/clarke-bui…
50537  http://www.tate.org.uk/art/artworks/clarke-pra…
53186  http://www.tate.org.uk/art/artworks/bourgeois-…
56283  http://www.tate.org.uk/art/artworks/hamilton-t…
67409  http://www.tate.org.uk/art/artworks/leach-2-st…
67432  http://www.tate.org.uk/art/artworks/leach-bowl…
```

```python
# strip whitespaces on entire series
data.title.str.strip()
```

```
0          A Figure Bowing before a Seated Old Man with h…
1          Two Drawings of Frightened Figures, Probably f…
2          The Preaching of Warning. Verso: An Old Man En…
3              Six Drawings of Figures with Outstretched Arms
4          The Circle of the Lustful: Francesca da Rimini…
                                …
69196                        Larvae (from Tampax Romana)
69197                    Living Womb (from Tampax Romana)
69198                                      Present Tense
69199             Work No. 227: The lights going on and off
69200                       Dancing Scene in the West Indies
Name: title, Length: 69201, dtype: object
```

```python
data.title = data.title.str.strip()
# str.lstrip()  # strip whitespace from left
# str.rstrip()  # strip whitespace from right
```

```python
data.loc[data.title.str.contains('\s$', regex=True)]
```

```
Empty DataFrame
Columns: [id, accession_number, artist, artistRole, artistId, title, dateText,
medium, creditLine, year, acquisitionYear, dimensions, width, height, depth,
units, inscription, thumbnailCopyright, thumbnailUrl, url]
```

```
Index: []
```

```
[ ]: data.title = data.title.transform(lambda x: x.strip())
```

```
[ ]: data.title
```

```
[ ]: 0        A Figure Bowing before a Seated Old Man with h…
     1        Two Drawings of Frightened Figures, Probably f…
     2        The Preaching of Warning. Verso: An Old Man En…
     3            Six Drawings of Figures with Outstretched Arms
     4        The Circle of the Lustful: Francesca da Rimini…
                                   …
     69196                       Larvae (from Tampax Romana)
     69197                   Living Womb (from Tampax Romana)
     69198                                      Present Tense
     69199         Work No. 227: The lights going on and off
     69200                     Dancing Scene in the West Indies
     Name: title, Length: 69201, dtype: object
```

1.0.2 Replacing Bad Data with NaN

```
[ ]: data = pd.read_csv('./data/artwork_data.csv', low_memory=False)
```

```
[ ]: data.head()
```

```
[ ]:      id accession_number         artist artistRole  artistId  \
     0  1035           A00001   Blake, Robert     artist        38
     1  1036           A00002   Blake, Robert     artist        38
     2  1037           A00003   Blake, Robert     artist        38
     3  1038           A00004   Blake, Robert     artist        38
     4  1039           A00005  Blake, William     artist        39

                                                    title              dateText  \
     0  A Figure Bowing before a Seated Old Man with h…         date not known
     1  Two Drawings of Frightened Figures, Probably f…         date not known
     2  The Preaching of Warning. Verso: An Old Man En…                ?c.1785
     3      Six Drawings of Figures with Outstretched Arms      date not known
     4  The Circle of the Lustful: Francesca da Rimini…  1826-7, reprinted 1892

                                          medium  \
     0  Watercolour, ink, chalk and graphite on paper…
     1                            Graphite on paper
     2         Graphite on paper. Verso: graphite on paper
     3                            Graphite on paper
     4                      Line engraving on paper

                                    creditLine  year  acquisitionYear  \
```

```
0                 Presented by Mrs John Richmond 1922    NaN           1922.0
1                 Presented by Mrs John Richmond 1922    NaN           1922.0
2                 Presented by Mrs John Richmond 1922   1785           1922.0
3                 Presented by Mrs John Richmond 1922    NaN           1922.0
4   Purchased with the assistance of a special gra…   1826           1919.0

              dimensions width height  depth units inscription  \
0  support: 394 x 419 mm    394    419    NaN    mm         NaN
1  support: 311 x 213 mm    311    213    NaN    mm         NaN
2  support: 343 x 467 mm    343    467    NaN    mm         NaN
3  support: 318 x 394 mm    318    394    NaN    mm         NaN
4    image: 243 x 335 mm    243    335    NaN    mm         NaN

  thumbnailCopyright                                       thumbnailUrl  \
0                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
1                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
2                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
3                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
4                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…

                                                 url
0  http://www.tate.org.uk/art/artworks/blake-a-fi…
1  http://www.tate.org.uk/art/artworks/blake-two-…
2  http://www.tate.org.uk/art/artworks/blake-the-…
3  http://www.tate.org.uk/art/artworks/blake-six-…
4  http://www.tate.org.uk/art/artworks/blake-the-…
```

```python
pd.isna(data.loc[:, 'dateText'])
```

```
0        False
1        False
2        False
3        False
4        False
         …
69196    False
69197    False
69198    False
69199    False
69200    False
Name: dateText, Length: 69201, dtype: bool
```

```python
import numpy as np
```

```python
data.replace({'dateText': {'date not known': np.nan}})
```

```
[ ]:             id accession_number            artist artistRole  artistId  \
      0         1035           A00001       Blake, Robert     artist        38
      1         1036           A00002       Blake, Robert     artist        38
      2         1037           A00003       Blake, Robert     artist        38
      3         1038           A00004       Blake, Robert     artist        38
      4         1039           A00005      Blake, William     artist        39
      …          …                …                 …          …         …
      69196   122960           T13865  P-Orridge, Genesis     artist     16646
      69197   122961           T13866  P-Orridge, Genesis     artist     16646
      69198   121181           T13867        Hatoum, Mona     artist      2365
      69199   112306           T13868        Creed, Martin     artist      2760
      69200   127035           T13869  Brunias, Agostino     artist      3824

                                                       title  \
      0      A Figure Bowing before a Seated Old Man with h…
      1      Two Drawings of Frightened Figures, Probably f…
      2      The Preaching of Warning. Verso: An Old Man En…
      3          Six Drawings of Figures with Outstretched Arms
      4      The Circle of the Lustful: Francesca da Rimini…
      …                                                    …
      69196                      Larvae (from Tampax Romana)
      69197                  Living Womb (from Tampax Romana)
      69198                                    Present Tense
      69199          Work No. 227: The lights going on and off
      69200                    Dancing Scene in the West Indies

                          dateText  \
      0                        NaN
      1                        NaN
      2                     ?c.1785
      3                        NaN
      4          1826-7, reprinted 1892
      …                          …
      69196                     1975
      69197                     1976
      69198                     1996
      69199                     2000
      69200                  1764-96

                                                      medium  \
      0      Watercolour, ink, chalk and graphite on paper…
      1                                   Graphite on paper
      2            Graphite on paper. Verso: graphite on paper
      3                                   Graphite on paper
      4                              Line engraving on paper
      …                                                    …
      69196   Perspex, Wood, hairpiece, tampon and human blood
```

```
69197  Wood, Perspex, plastic, photograph on paper, t…
69198                          Soap and glass beads
69199                               Gallery lighting
69200                             Oil paint on canvas


                                      creditLine  year  \
0                 Presented by Mrs John Richmond 1922   NaN
1                 Presented by Mrs John Richmond 1922   NaN
2                 Presented by Mrs John Richmond 1922   1785
3                 Presented by Mrs John Richmond 1922   NaN
4       Purchased with the assistance of a special gra…  1826
…                                              …    …
69196              Transferred from Tate Archive 2012  1975
69197              Transferred from Tate Archive 2012  1976
69198               Presented by Tate Members 2013  1996
69199   Purchased with funds provided by Tate Members,…  2000
69200   Purchased with assistance from Tate Patrons an…  1764


       acquisitionYear                        dimensions width height  \
0               1922.0            support: 394 x 419 mm    394    419
1               1922.0            support: 311 x 213 mm    311    213
2               1922.0            support: 343 x 467 mm    343    467
3               1922.0            support: 318 x 394 mm    318    394
4               1919.0              image: 243 x 335 mm    243    335
…                   …                             …    …     …
69196           2013.0        object: 305 x 305 x 135 mm   305    305
69197           2013.0        object: 305 x 305 x 135 mm   305    305
69198           2013.0      displayed: 45 x 2410 x 2990 mm    45   2410
69199           2013.0  Overall display dimensions variable  NaN    NaN
69200           2013.0           unconfirmed: 508 x 660 mm   508    660


       depth units inscription thumbnailCopyright  \
0        NaN   mm         NaN              NaN
1        NaN   mm         NaN              NaN
2        NaN   mm         NaN              NaN
3        NaN   mm         NaN              NaN
4        NaN   mm         NaN              NaN
…         …    …           …                …
69196  135.0   mm         NaN              NaN
69197  135.0   mm         NaN              NaN
69198 2990.0   mm         NaN              NaN
69199    NaN  NaN         NaN              NaN
69200    NaN   mm         NaN              NaN


                                    thumbnailUrl  \
0      http://www.tate.org.uk/art/images/work/A/A00/A…
1      http://www.tate.org.uk/art/images/work/A/A00/A…
```

```
2      http://www.tate.org.uk/art/images/work/A/A00/A…
3      http://www.tate.org.uk/art/images/work/A/A00/A…
4      http://www.tate.org.uk/art/images/work/A/A00/A…
…                                                    …
69196                                              NaN
69197                                              NaN
69198                                              NaN
69199                                              NaN
69200  http://www.tate.org.uk/art/images/work/T/T13/T…

                                                    url
0      http://www.tate.org.uk/art/artworks/blake-a-fi…
1      http://www.tate.org.uk/art/artworks/blake-two-…
2      http://www.tate.org.uk/art/artworks/blake-the-…
3      http://www.tate.org.uk/art/artworks/blake-six-…
4      http://www.tate.org.uk/art/artworks/blake-the-…
…                                                    …
69196  http://www.tate.org.uk/art/artworks/p-orridge-…
69197  http://www.tate.org.uk/art/artworks/p-orridge-…
69198  http://www.tate.org.uk/art/artworks/hatoum-pre…
69199  http://www.tate.org.uk/art/artworks/creed-work…
69200  http://www.tate.org.uk/art/artworks/brunias-da…

[69201 rows x 20 columns]
```

`[ ]:` `data.head()`

```
[ ]:       id accession_number          artist artistRole  artistId  \
     0   1035           A00001   Blake, Robert     artist        38
     1   1036           A00002   Blake, Robert     artist        38
     2   1037           A00003   Blake, Robert     artist        38
     3   1038           A00004   Blake, Robert     artist        38
     4   1039           A00005  Blake, William     artist        39

                                                    title              dateText  \
     0  A Figure Bowing before a Seated Old Man with h…       date not known
     1  Two Drawings of Frightened Figures, Probably f…       date not known
     2  The Preaching of Warning. Verso: An Old Man En…               ?c.1785
     3     Six Drawings of Figures with Outstretched Arms      date not known
     4  The Circle of the Lustful: Francesca da Rimini…  1826-7, reprinted 1892

                                                    medium  \
     0  Watercolour, ink, chalk and graphite on paper…
     1                                Graphite on paper
     2        Graphite on paper. Verso: graphite on paper
     3                                Graphite on paper
     4                        Line engraving on paper
```

```
                                      creditLine  year  acquisitionYear  \
0                Presented by Mrs John Richmond 1922   NaN           1922.0
1                Presented by Mrs John Richmond 1922   NaN           1922.0
2                Presented by Mrs John Richmond 1922  1785           1922.0
3                Presented by Mrs John Richmond 1922   NaN           1922.0
4  Purchased with the assistance of a special gra… 1826           1919.0

              dimensions width height  depth units inscription  \
0  support: 394 x 419 mm    394    419   NaN    mm         NaN
1  support: 311 x 213 mm    311    213   NaN    mm         NaN
2  support: 343 x 467 mm    343    467   NaN    mm         NaN
3  support: 318 x 394 mm    318    394   NaN    mm         NaN
4    image: 243 x 335 mm    243    335   NaN    mm         NaN

  thumbnailCopyright                                       thumbnailUrl  \
0                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
1                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
2                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
3                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
4                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…

                                                 url
0  http://www.tate.org.uk/art/artworks/blake-a-fi…
1  http://www.tate.org.uk/art/artworks/blake-two-…
2  http://www.tate.org.uk/art/artworks/blake-the-…
3  http://www.tate.org.uk/art/artworks/blake-six-…
4  http://www.tate.org.uk/art/artworks/blake-the-…
```

```python
data.replace({'dateText': {'date not known': np.nan}}, inplace=True)
```

```python
data.head()
```

```
     id accession_number          artist artistRole  artistId  \
0  1035           A00001   Blake, Robert     artist        38
1  1036           A00002   Blake, Robert     artist        38
2  1037           A00003   Blake, Robert     artist        38
3  1038           A00004   Blake, Robert     artist        38
4  1039           A00005  Blake, William     artist        39

                                               title              dateText  \
0  A Figure Bowing before a Seated Old Man with h…                   NaN
1  Two Drawings of Frightened Figures, Probably f…                   NaN
2  The Preaching of Warning. Verso: An Old Man En…                ?c.1785
3      Six Drawings of Figures with Outstretched Arms                 NaN
4  The Circle of the Lustful: Francesca da Rimini…  1826-7, reprinted 1892
```

```
                                                medium  \
0   Watercolour, ink, chalk and graphite on paper…
1                                  Graphite on paper
2        Graphite on paper. Verso: graphite on paper
3                                  Graphite on paper
4                            Line engraving on paper


                                     creditLine  year  acquisitionYear  \
0              Presented by Mrs John Richmond 1922   NaN           1922.0
1              Presented by Mrs John Richmond 1922   NaN           1922.0
2              Presented by Mrs John Richmond 1922  1785           1922.0
3              Presented by Mrs John Richmond 1922   NaN           1922.0
4   Purchased with the assistance of a special gra…  1826           1919.0


              dimensions  width  height  depth  units  inscription  \
0   support: 394 x 419 mm    394     419    NaN     mm          NaN
1   support: 311 x 213 mm    311     213    NaN     mm          NaN
2   support: 343 x 467 mm    343     467    NaN     mm          NaN
3   support: 318 x 394 mm    318     394    NaN     mm          NaN
4     image: 243 x 335 mm    243     335    NaN     mm          NaN


  thumbnailCopyright                                        thumbnailUrl  \
0                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
1                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
2                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
3                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
4                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…


                                                url
0   http://www.tate.org.uk/art/artworks/blake-a-fi…
1   http://www.tate.org.uk/art/artworks/blake-two-…
2   http://www.tate.org.uk/art/artworks/blake-the-…
3   http://www.tate.org.uk/art/artworks/blake-six-…
4   http://www.tate.org.uk/art/artworks/blake-the-…
```

```python
# reload data
data = pd.read_csv('./data/artwork_data.csv', low_memory=False)
```

```python
data.head()
```

```
     id accession_number          artist artistRole  artistId  \
0  1035           A00001   Blake, Robert     artist        38
1  1036           A00002   Blake, Robert     artist        38
2  1037           A00003   Blake, Robert     artist        38
3  1038           A00004   Blake, Robert     artist        38
4  1039           A00005  Blake, William     artist        39
```

```
                                                  title               dateText  \
0  A Figure Bowing before a Seated Old Man with h…          date not known
1  Two Drawings of Frightened Figures, Probably f…          date not known
2  The Preaching of Warning. Verso: An Old Man En…                  ?c.1785
3     Six Drawings of Figures with Outstretched Arms          date not known
4  The Circle of the Lustful: Francesca da Rimini…  1826-7, reprinted 1892

                                            medium  \
0  Watercolour, ink, chalk and graphite on paper…
1                                Graphite on paper
2        Graphite on paper. Verso: graphite on paper
3                                Graphite on paper
4                          Line engraving on paper

                                        creditLine  year  acquisitionYear  \
0              Presented by Mrs John Richmond 1922   NaN           1922.0
1              Presented by Mrs John Richmond 1922   NaN           1922.0
2              Presented by Mrs John Richmond 1922  1785           1922.0
3              Presented by Mrs John Richmond 1922   NaN           1922.0
4  Purchased with the assistance of a special gra…  1826           1919.0

              dimensions width height  depth units inscription  \
0  support: 394 x 419 mm    394    419    NaN    mm         NaN
1  support: 311 x 213 mm    311    213    NaN    mm         NaN
2  support: 343 x 467 mm    343    467    NaN    mm         NaN
3  support: 318 x 394 mm    318    394    NaN    mm         NaN
4    image: 243 x 335 mm    243    335    NaN    mm         NaN

  thumbnailCopyright                                       thumbnailUrl  \
0                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
1                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
2                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
3                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
4                NaN  http://www.tate.org.uk/art/images/work/A/A00/A…

                                                 url
0  http://www.tate.org.uk/art/artworks/blake-a-fi…
1  http://www.tate.org.uk/art/artworks/blake-two-…
2  http://www.tate.org.uk/art/artworks/blake-the-…
3  http://www.tate.org.uk/art/artworks/blake-six-…
4  http://www.tate.org.uk/art/artworks/blake-the-…
```

```python
# filter data and then replace it with NaN
data.loc[data.dateText == 'date not known', ['dateText']] = np.nan
```

```python
data.head()
```

```
[ ]:      id accession_number         artist artistRole  artistId  \
    0  1035           A00001   Blake, Robert     artist        38
    1  1036           A00002   Blake, Robert     artist        38
    2  1037           A00003   Blake, Robert     artist        38
    3  1038           A00004   Blake, Robert     artist        38
    4  1039           A00005  Blake, William     artist        39

                                            title           dateText  \
    0  A Figure Bowing before a Seated Old Man with h…                 NaN
    1  Two Drawings of Frightened Figures, Probably f…                 NaN
    2  The Preaching of Warning. Verso: An Old Man En…             ?c.1785
    3     Six Drawings of Figures with Outstretched Arms              NaN
    4  The Circle of the Lustful: Francesca da Rimini…  1826-7, reprinted 1892

                                           medium  \
    0  Watercolour, ink, chalk and graphite on paper…
    1                             Graphite on paper
    2        Graphite on paper. Verso: graphite on paper
    3                             Graphite on paper
    4                       Line engraving on paper

                                          creditLine  year  acquisitionYear  \
    0              Presented by Mrs John Richmond 1922   NaN           1922.0
    1              Presented by Mrs John Richmond 1922   NaN           1922.0
    2              Presented by Mrs John Richmond 1922  1785           1922.0
    3              Presented by Mrs John Richmond 1922   NaN           1922.0
    4  Purchased with the assistance of a special gra…  1826           1919.0

                dimensions  width height  depth units inscription  \
    0  support: 394 x 419 mm    394    419    NaN    mm         NaN
    1  support: 311 x 213 mm    311    213    NaN    mm         NaN
    2  support: 343 x 467 mm    343    467    NaN    mm         NaN
    3  support: 318 x 394 mm    318    394    NaN    mm         NaN
    4    image: 243 x 335 mm    243    335    NaN    mm         NaN

       thumbnailCopyright                                        thumbnailUrl  \
    0                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
    1                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
    2                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
    3                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
    4                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…

                                         url
    0  http://www.tate.org.uk/art/artworks/blake-a-fi…
    1  http://www.tate.org.uk/art/artworks/blake-two-…
    2  http://www.tate.org.uk/art/artworks/blake-the-…
    3  http://www.tate.org.uk/art/artworks/blake-six-…
```

```
4    http://www.tate.org.uk/art/artworks/blake-the-…
```

```
[ ]: data.loc[data.year.notnull() & data.year.astype(str).str.contains('[^0-9]')]
```

```
[ ]:            id accession_number           artist artistRole  artistId  \
     67968    99332           T12629  Roberts, William     artist      1855
     67980    99346           T12641  Roberts, William     artist      1855
     67987    99354           T12648  Roberts, William     artist      1855
     67994    99361           T12655  Roberts, William     artist      1855
     68002    99370           T12663  Roberts, William     artist      1855
     68015    99384           T12676  Roberts, William     artist      1855
     68033    99407           T12694  Roberts, William     artist      1855
     68042    99426           T12703  Roberts, William     artist      1855
     68047    99431           T12708  Roberts, William     artist      1855
     68051    99435           T12712  Roberts, William     artist      1855
     68056    99440           T12717  Roberts, William     artist      1855
     68057    99441           T12718  Roberts, William     artist      1855
     68062    99446           T12723  Roberts, William     artist      1855
     68070    99454           T12731  Roberts, William     artist      1855
     68072    99456           T12733  Roberts, William     artist      1855
     68074    99458           T12735  Roberts, William     artist      1855
     68076    99460           T12737  Roberts, William     artist      1855
     68077    99462           T12738  Roberts, William     artist      1855
     69165   121190           T13834  Barlow, Phyllida     artist     10908

                                                      title  dateText  \
     67968                              Families on a beach   no date
     67980                             Peasants and horseman   no date
     67987                                       The Beatles   no date
     67994                                       The peasants   no date
     68002                                 Two ladies talking   no date
     68015                                     Butcher's Man   no date
     68033                                       Men fishing   no date
     68042                                         Bar scene   no date
     68047  The critic intervenes; Man milking a goat (rev…   no date
     68051                      Woman brushing another's hair   no date
     68056                             Group of people in a field   no date
     68057                        Two women playing with a cat   no date
     68062              Group of people standing on a corner   no date
     68070                             Group of people in a punt   no date
     68072                                         Party scene   no date
     68074                                       Lady in a hat   no date
     68076                                           The toast   no date
     68077                  People diving into a swimming pool   no date
     69165                                           Untitled  c.1997-9

                      medium  \
```

```
67968               Graphite on paper
67980  Graphite and watercolour on paper
67987               Graphite on paper
67994               Graphite on paper
68002               Graphite on paper
68015               Graphite on paper
68033               Graphite on paper
68042               Graphite on paper
68047               Graphite on paper
68051  Watercolour and graphite on paper
68056               Graphite on paper
68057               Graphite on paper
68062               Graphite on paper
68070               Graphite on paper
68072               Graphite on paper
68074               Graphite on paper
68076               Graphite on paper
68077               Graphite on paper
69165            Acrylic paint on paper

                                          creditLine       year  \
67968  Accepted by HM Government in lieu of inheritan…   no date
67980  Accepted by HM Government in lieu of inheritan…   no date
67987  Accepted by HM Government in lieu of inheritan…   no date
67994  Accepted by HM Government in lieu of inheritan…   no date
68002  Accepted by HM Government in lieu of inheritan…   no date
68015  Accepted by HM Government in lieu of inheritan…   no date
68033  Accepted by HM Government in lieu of inheritan…   no date
68042  Accepted by HM Government in lieu of inheritan…   no date
68047  Accepted by HM Government in lieu of inheritan…   no date
68051  Accepted by HM Government in lieu of inheritan…   no date
68056  Accepted by HM Government in lieu of inheritan…   no date
68057  Accepted by HM Government in lieu of inheritan…   no date
68062  Accepted by HM Government in lieu of inheritan…   no date
68070  Accepted by HM Government in lieu of inheritan…   no date
68072  Accepted by HM Government in lieu of inheritan…   no date
68074  Accepted by HM Government in lieu of inheritan…   no date
68076  Accepted by HM Government in lieu of inheritan…   no date
68077  Accepted by HM Government in lieu of inheritan…   no date
69165  Presented by the Trustees of the Chantrey Bequ…  c.1997-9

      acquisitionYear              dimensions width height  depth units  \
67968          2008.0     support: 127 x 178 mm   127    178    NaN    mm
67980          2008.0             184 x 140 mm   NaN    NaN    NaN   NaN
67987          2008.0             178 x 127 mm   NaN    NaN    NaN   NaN
67994          2008.0             191 x 133 mm   NaN    NaN    NaN   NaN
68002          2008.0             178 x 127 mm   NaN    NaN    NaN   NaN
```

68015	2008.0	support: 267 x 191 mm	267	191	NaN	mm
68033	2008.0	203 x 172 mm	NaN	NaN	NaN	NaN
68042	2008.0	127 x 184 mm	NaN	NaN	NaN	NaN
68047	2008.0	152 x 133 mm	NaN	NaN	NaN	NaN
68051	2008.0	178 x 89 mm	NaN	NaN	NaN	NaN
68056	2008.0	187 x 133 mm	NaN	NaN	NaN	NaN
68057	2008.0	1528 x 121 mm	NaN	NaN	NaN	NaN
68062	2008.0	184 x 133 mm	NaN	NaN	NaN	NaN
68070	2008.0	273 x 172 mm	NaN	NaN	NaN	NaN
68072	2008.0	122 x 108 mm	NaN	NaN	NaN	NaN
68074	2008.0	184 x 133 mm	NaN	NaN	NaN	NaN
68076	2008.0	178 x 121 mm	NaN	NaN	NaN	NaN
68077	2008.0	184 x 162 mm	NaN	NaN	NaN	NaN
69165	2013.0	unconfirmed: 389 x 562 mm	389	562	NaN	mm

	inscription	thumbnailCopyright \
67968	NaN	© The estate of William Roberts
67980	NaN	© The estate of William Roberts
67987	NaN	© The estate of William Roberts
67994	NaN	© The estate of William Roberts
68002	NaN	© The estate of William Roberts
68015	NaN	© The estate of William Roberts
68033	NaN	© The estate of William Roberts
68042	NaN	© The estate of William Roberts
68047	NaN	© The estate of William Roberts
68051	NaN	© The estate of William Roberts
68056	NaN	© The estate of William Roberts
68057	NaN	© The estate of William Roberts
68062	NaN	© The estate of William Roberts
68070	NaN	NaN
68072	NaN	© The estate of William Roberts
68074	NaN	© The estate of William Roberts
68076	NaN	© The estate of William Roberts
68077	NaN	© The estate of William Roberts
69165	NaN	© Phyllida Barlow

	thumbnailUrl \
67968	http://www.tate.org.uk/art/images/work/T/T12/T…
67980	http://www.tate.org.uk/art/images/work/T/T12/T…
67987	http://www.tate.org.uk/art/images/work/T/T12/T…
67994	http://www.tate.org.uk/art/images/work/T/T12/T…
68002	http://www.tate.org.uk/art/images/work/T/T12/T…
68015	http://www.tate.org.uk/art/images/work/T/T12/T…
68033	http://www.tate.org.uk/art/images/work/T/T12/T…
68042	http://www.tate.org.uk/art/images/work/T/T12/T…
68047	http://www.tate.org.uk/art/images/work/T/T12/T…
68051	http://www.tate.org.uk/art/images/work/T/T12/T…

```
68056  http://www.tate.org.uk/art/images/work/T/T12/T…
68057  http://www.tate.org.uk/art/images/work/T/T12/T…
68062  http://www.tate.org.uk/art/images/work/T/T12/T…
68070                                             NaN
68072  http://www.tate.org.uk/art/images/work/T/T12/T…
68074  http://www.tate.org.uk/art/images/work/T/T12/T…
68076  http://www.tate.org.uk/art/images/work/T/T12/T…
68077  http://www.tate.org.uk/art/images/work/T/T12/T…
69165  http://www.tate.org.uk/art/images/work/T/T13/T…

                                                   url
67968  http://www.tate.org.uk/art/artworks/roberts-fa…
67980  http://www.tate.org.uk/art/artworks/roberts-pe…
67987  http://www.tate.org.uk/art/artworks/roberts-th…
67994  http://www.tate.org.uk/art/artworks/roberts-th…
68002  http://www.tate.org.uk/art/artworks/roberts-tw…
68015  http://www.tate.org.uk/art/artworks/roberts-bu…
68033  http://www.tate.org.uk/art/artworks/roberts-me…
68042  http://www.tate.org.uk/art/artworks/roberts-ba…
68047  http://www.tate.org.uk/art/artworks/roberts-th…
68051  http://www.tate.org.uk/art/artworks/roberts-wo…
68056  http://www.tate.org.uk/art/artworks/roberts-gr…
68057  http://www.tate.org.uk/art/artworks/roberts-tw…
68062  http://www.tate.org.uk/art/artworks/roberts-gr…
68070  http://www.tate.org.uk/art/artworks/roberts-gr…
68072  http://www.tate.org.uk/art/artworks/roberts-pa…
68074  http://www.tate.org.uk/art/artworks/roberts-la…
68076  http://www.tate.org.uk/art/artworks/roberts-th…
68077  http://www.tate.org.uk/art/artworks/roberts-pe…
69165  http://www.tate.org.uk/art/artworks/barlow-unt…
```

```python
data.year.notnull()
```

```
0        False
1        False
2         True
3        False
4         True
         ...
69196     True
69197     True
69198     True
69199     True
69200     True
Name: year, Length: 69201, dtype: bool
```

```python
data.year.astype(str).str.contains('[^0-9]')
```

```
[ ]: 0          True
     1          True
     2         False
     3          True
     4         False
                 …
     69196     False
     69197     False
     69198     False
     69199     False
     69200     False
     Name: year, Length: 69201, dtype: bool
```

```
[ ]: data.year.notnull() & data.year.astype(str).str.contains('[^0-9]')
```

```
[ ]: 0         False
     1         False
     2         False
     3         False
     4         False
                 …
     69196     False
     69197     False
     69198     False
     69199     False
     69200     False
     Name: year, Length: 69201, dtype: bool
```

```
[ ]: data.loc[data.year.notnull() & data.year.astype(str).str.contains('[^0-9]')]
```

```
[ ]:         id accession_number            artist artistRole  artistId  \
     67968  99332           T12629  Roberts, William    artist      1855
     67980  99346           T12641  Roberts, William    artist      1855
     67987  99354           T12648  Roberts, William    artist      1855
     67994  99361           T12655  Roberts, William    artist      1855
     68002  99370           T12663  Roberts, William    artist      1855
     68015  99384           T12676  Roberts, William    artist      1855
     68033  99407           T12694  Roberts, William    artist      1855
     68042  99426           T12703  Roberts, William    artist      1855
     68047  99431           T12708  Roberts, William    artist      1855
     68051  99435           T12712  Roberts, William    artist      1855
     68056  99440           T12717  Roberts, William    artist      1855
     68057  99441           T12718  Roberts, William    artist      1855
     68062  99446           T12723  Roberts, William    artist      1855
     68070  99454           T12731  Roberts, William    artist      1855
     68072  99456           T12733  Roberts, William    artist      1855
     68074  99458           T12735  Roberts, William    artist      1855
```

```
68076    99460              T12737  Roberts, William    artist      1855
68077    99462              T12738  Roberts, William    artist      1855
69165   121190              T13834  Barlow, Phyllida    artist     10908
```

```
                                                    title  dateText  \
67968                               Families on a beach   no date
67980                             Peasants and horseman   no date
67987                                       The Beatles   no date
67994                                       The peasants   no date
68002                                 Two ladies talking   no date
68015                                       Butcher's Man   no date
68033                                       Men fishing   no date
68042                                         Bar scene   no date
68047  The critic intervenes; Man milking a goat (rev…   no date
68051                     Woman brushing another's hair   no date
68056                           Group of people in a field   no date
68057                         Two women playing with a cat   no date
68062                   Group of people standing on a corner   no date
68070                           Group of people in a punt   no date
68072                                       Party scene   no date
68074                                       Lady in a hat   no date
68076                                         The toast   no date
68077                     People diving into a swimming pool   no date
69165                                           Untitled  c.1997-9
```

```
                                medium  \
67968                Graphite on paper
67980  Graphite and watercolour on paper
67987                Graphite on paper
67994                Graphite on paper
68002                Graphite on paper
68015                Graphite on paper
68033                Graphite on paper
68042                Graphite on paper
68047                Graphite on paper
68051  Watercolour and graphite on paper
68056                Graphite on paper
68057                Graphite on paper
68062                Graphite on paper
68070                Graphite on paper
68072                Graphite on paper
68074                Graphite on paper
68076                Graphite on paper
68077                Graphite on paper
69165          Acrylic paint on paper
```

```
                            creditLine      year  \
```

```
67968  Accepted by HM Government in lieu of inheritan…   no date
67980  Accepted by HM Government in lieu of inheritan…   no date
67987  Accepted by HM Government in lieu of inheritan…   no date
67994  Accepted by HM Government in lieu of inheritan…   no date
68002  Accepted by HM Government in lieu of inheritan…   no date
68015  Accepted by HM Government in lieu of inheritan…   no date
68033  Accepted by HM Government in lieu of inheritan…   no date
68042  Accepted by HM Government in lieu of inheritan…   no date
68047  Accepted by HM Government in lieu of inheritan…   no date
68051  Accepted by HM Government in lieu of inheritan…   no date
68056  Accepted by HM Government in lieu of inheritan…   no date
68057  Accepted by HM Government in lieu of inheritan…   no date
68062  Accepted by HM Government in lieu of inheritan…   no date
68070  Accepted by HM Government in lieu of inheritan…   no date
68072  Accepted by HM Government in lieu of inheritan…   no date
68074  Accepted by HM Government in lieu of inheritan…   no date
68076  Accepted by HM Government in lieu of inheritan…   no date
68077  Accepted by HM Government in lieu of inheritan…   no date
69165  Presented by the Trustees of the Chantrey Bequ…  c.1997-9

       acquisitionYear                    dimensions width height  depth units  \
67968           2008.0     support: 127 x 178 mm     127    178    NaN    mm
67980           2008.0              184 x 140 mm     NaN    NaN    NaN   NaN
67987           2008.0              178 x 127 mm     NaN    NaN    NaN   NaN
67994           2008.0              191 x 133 mm     NaN    NaN    NaN   NaN
68002           2008.0              178 x 127 mm     NaN    NaN    NaN   NaN
68015           2008.0     support: 267 x 191 mm     267    191    NaN    mm
68033           2008.0              203 x 172 mm     NaN    NaN    NaN   NaN
68042           2008.0              127 x 184 mm     NaN    NaN    NaN   NaN
68047           2008.0              152 x 133 mm     NaN    NaN    NaN   NaN
68051           2008.0               178 x 89 mm     NaN    NaN    NaN   NaN
68056           2008.0              187 x 133 mm     NaN    NaN    NaN   NaN
68057           2008.0             1528 x 121 mm     NaN    NaN    NaN   NaN
68062           2008.0              184 x 133 mm     NaN    NaN    NaN   NaN
68070           2008.0              273 x 172 mm     NaN    NaN    NaN   NaN
68072           2008.0              122 x 108 mm     NaN    NaN    NaN   NaN
68074           2008.0              184 x 133 mm     NaN    NaN    NaN   NaN
68076           2008.0              178 x 121 mm     NaN    NaN    NaN   NaN
68077           2008.0              184 x 162 mm     NaN    NaN    NaN   NaN
69165           2013.0  unconfirmed: 389 x 562 mm     389    562    NaN    mm

        inscription              thumbnailCopyright  \
67968           NaN  © The estate of William Roberts
67980           NaN  © The estate of William Roberts
67987           NaN  © The estate of William Roberts
67994           NaN  © The estate of William Roberts
68002           NaN  © The estate of William Roberts
```

```
68015           NaN  © The estate of William Roberts
68033           NaN  © The estate of William Roberts
68042           NaN  © The estate of William Roberts
68047           NaN  © The estate of William Roberts
68051           NaN  © The estate of William Roberts
68056           NaN  © The estate of William Roberts
68057           NaN  © The estate of William Roberts
68062           NaN  © The estate of William Roberts
68070           NaN                              NaN
68072           NaN  © The estate of William Roberts
68074           NaN  © The estate of William Roberts
68076           NaN  © The estate of William Roberts
68077           NaN  © The estate of William Roberts
69165           NaN              © Phyllida Barlow


                                     thumbnailUrl  \
67968  http://www.tate.org.uk/art/images/work/T/T12/T…
67980  http://www.tate.org.uk/art/images/work/T/T12/T…
67987  http://www.tate.org.uk/art/images/work/T/T12/T…
67994  http://www.tate.org.uk/art/images/work/T/T12/T…
68002  http://www.tate.org.uk/art/images/work/T/T12/T…
68015  http://www.tate.org.uk/art/images/work/T/T12/T…
68033  http://www.tate.org.uk/art/images/work/T/T12/T…
68042  http://www.tate.org.uk/art/images/work/T/T12/T…
68047  http://www.tate.org.uk/art/images/work/T/T12/T…
68051  http://www.tate.org.uk/art/images/work/T/T12/T…
68056  http://www.tate.org.uk/art/images/work/T/T12/T…
68057  http://www.tate.org.uk/art/images/work/T/T12/T…
68062  http://www.tate.org.uk/art/images/work/T/T12/T…
68070                                           NaN
68072  http://www.tate.org.uk/art/images/work/T/T12/T…
68074  http://www.tate.org.uk/art/images/work/T/T12/T…
68076  http://www.tate.org.uk/art/images/work/T/T12/T…
68077  http://www.tate.org.uk/art/images/work/T/T12/T…
69165  http://www.tate.org.uk/art/images/work/T/T13/T…


                                              url
67968  http://www.tate.org.uk/art/artworks/roberts-fa…
67980  http://www.tate.org.uk/art/artworks/roberts-pe…
67987  http://www.tate.org.uk/art/artworks/roberts-th…
67994  http://www.tate.org.uk/art/artworks/roberts-th…
68002  http://www.tate.org.uk/art/artworks/roberts-tw…
68015  http://www.tate.org.uk/art/artworks/roberts-bu…
68033  http://www.tate.org.uk/art/artworks/roberts-me…
68042  http://www.tate.org.uk/art/artworks/roberts-ba…
68047  http://www.tate.org.uk/art/artworks/roberts-th…
68051  http://www.tate.org.uk/art/artworks/roberts-wo…
```

```
68056    http://www.tate.org.uk/art/artworks/roberts-gr…
68057    http://www.tate.org.uk/art/artworks/roberts-tw…
68062    http://www.tate.org.uk/art/artworks/roberts-gr…
68070    http://www.tate.org.uk/art/artworks/roberts-gr…
68072    http://www.tate.org.uk/art/artworks/roberts-pa…
68074    http://www.tate.org.uk/art/artworks/roberts-la…
68076    http://www.tate.org.uk/art/artworks/roberts-th…
68077    http://www.tate.org.uk/art/artworks/roberts-pe…
69165    http://www.tate.org.uk/art/artworks/barlow-unt…
```

```python
data.loc[data.year.notnull() & data.year.astype(str).str.contains('[^0-9]'),
    ['year']] = np.nan
```

```python
data.loc[data.year.notnull() & data.year.astype(str).str.contains('[^0-9]'),
    ['year']]
```

```
Empty DataFrame
Columns: [year]
Index: []
```

```python
year_index = list(data.columns).index('year')
data.iloc[68077:68078, year_index: year_index+1]
```

```
        year
68077   NaN
```

```python
# reload the data
data = pd.read_csv('./data/artwork_data.csv', low_memory=False)
```

```python
data.head()
```

```
      id accession_number        artist artistRole  artistId  \
0   1035          A00001   Blake, Robert     artist        38
1   1036          A00002   Blake, Robert     artist        38
2   1037          A00003   Blake, Robert     artist        38
3   1038          A00004   Blake, Robert     artist        38
4   1039          A00005  Blake, William     artist        39

                                            title              dateText  \
0  A Figure Bowing before a Seated Old Man with h…        date not known
1  Two Drawings of Frightened Figures, Probably f…        date not known
2  The Preaching of Warning. Verso: An Old Man En…                 ?c.1785
3     Six Drawings of Figures with Outstretched Arms        date not known
4  The Circle of the Lustful: Francesca da Rimini…  1826-7, reprinted 1892

                                       medium  \
0  Watercolour, ink, chalk and graphite on paper…
1                            Graphite on paper
```

```
2        Graphite on paper. Verso: graphite on paper
3                        Graphite on paper
4                    Line engraving on paper


                                      creditLine  year  acquisitionYear  \
0            Presented by Mrs John Richmond 1922   NaN           1922.0
1            Presented by Mrs John Richmond 1922   NaN           1922.0
2            Presented by Mrs John Richmond 1922  1785           1922.0
3            Presented by Mrs John Richmond 1922   NaN           1922.0
4  Purchased with the assistance of a special gra…  1826           1919.0


              dimensions width height  depth units inscription  \
0  support: 394 x 419 mm    394    419   NaN    mm         NaN
1  support: 311 x 213 mm    311    213   NaN    mm         NaN
2  support: 343 x 467 mm    343    467   NaN    mm         NaN
3  support: 318 x 394 mm    318    394   NaN    mm         NaN
4    image: 243 x 335 mm    243    335   NaN    mm         NaN


  thumbnailCopyright                               thumbnailUrl  \
0               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
1               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
2               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
3               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
4               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…


                                            url
0  http://www.tate.org.uk/art/artworks/blake-a-fi…
1  http://www.tate.org.uk/art/artworks/blake-two-…
2  http://www.tate.org.uk/art/artworks/blake-the-…
3  http://www.tate.org.uk/art/artworks/blake-six-…
4  http://www.tate.org.uk/art/artworks/blake-the-…
```

```python
data[['year', 'depth']]
```

```
       year   depth
0       NaN     NaN
1       NaN     NaN
2      1785     NaN
3       NaN     NaN
4      1826     NaN
…       …       …
69196  1975   135.0
69197  1976   135.0
69198  1996  2990.0
69199  2000     NaN
69200  1764     NaN
```

```
[69201 rows x 2 columns]
```

[ ]: `data.fillna(0)`

[ ]:
```
              id accession_number             artist artistRole  artistId  \
0           1035          A00001      Blake, Robert     artist        38
1           1036          A00002      Blake, Robert     artist        38
2           1037          A00003      Blake, Robert     artist        38
3           1038          A00004      Blake, Robert     artist        38
4           1039          A00005    Blake, William     artist        39
...          ...             ...                ...        ...       ...
69196     122960          T13865  P-Orridge, Genesis     artist     16646
69197     122961          T13866  P-Orridge, Genesis     artist     16646
69198     121181          T13867       Hatoum, Mona     artist      2365
69199     112306          T13868       Creed, Martin     artist      2760
69200     127035          T13869  Brunias, Agostino     artist      3824

                                                    title  \
0          A Figure Bowing before a Seated Old Man with h…
1          Two Drawings of Frightened Figures, Probably f…
2          The Preaching of Warning. Verso: An Old Man En…
3            Six Drawings of Figures with Outstretched Arms
4          The Circle of the Lustful: Francesca da Rimini…
...                                                     …
69196                        Larvae (from Tampax Romana)
69197                    Living Womb (from Tampax Romana)
69198                                      Present Tense
69199            Work No. 227: The lights going on and off
69200                       Dancing Scene in the West Indies

                    dateText  \
0                date not known
1                date not known
2                       ?c.1785
3                date not known
4          1826-7, reprinted 1892
...                          …
69196                      1975
69197                      1976
69198                      1996
69199                      2000
69200                   1764-96

                                                   medium  \
0          Watercolour, ink, chalk and graphite on paper…
1                                     Graphite on paper
2                Graphite on paper. Verso: graphite on paper
```

```
3                                  Graphite on paper
4                            Line engraving on paper
...                                               ...
69196   Perspex, Wood, hairpiece, tampon and human blood
69197   Wood, Perspex, plastic, photograph on paper, t…
69198                             Soap and glass beads
69199                                  Gallery lighting
69200                               Oil paint on canvas

                                                 creditLine  year  \
0                    Presented by Mrs John Richmond 1922      0
1                    Presented by Mrs John Richmond 1922      0
2                    Presented by Mrs John Richmond 1922   1785
3                    Presented by Mrs John Richmond 1922      0
4          Purchased with the assistance of a special gra…  1826
...                                                    ...   ...
69196                  Transferred from Tate Archive 2012   1975
69197                  Transferred from Tate Archive 2012   1976
69198                    Presented by Tate Members 2013   1996
69199      Purchased with funds provided by Tate Members,…  2000
69200      Purchased with assistance from Tate Patrons an…  1764

       acquisitionYear                       dimensions width height  \
0                 1922.0            support: 394 x 419 mm    394    419
1                 1922.0            support: 311 x 213 mm    311    213
2                 1922.0            support: 343 x 467 mm    343    467
3                 1922.0            support: 318 x 394 mm    318    394
4                 1919.0               image: 243 x 335 mm    243    335
...                   ...                              ...    ...    ...
69196             2013.0        object: 305 x 305 x 135 mm    305    305
69197             2013.0        object: 305 x 305 x 135 mm    305    305
69198             2013.0      displayed: 45 x 2410 x 2990 mm    45   2410
69199             2013.0  Overall display dimensions variable      0      0
69200             2013.0            unconfirmed: 508 x 660 mm    508    660

        depth units inscription thumbnailCopyright  \
0         0.0    mm           0                  0
1         0.0    mm           0                  0
2         0.0    mm           0                  0
3         0.0    mm           0                  0
4         0.0    mm           0                  0
...       ...   ...          ...                ...
69196   135.0    mm           0                  0
69197   135.0    mm           0                  0
69198  2990.0    mm           0                  0
69199     0.0     0           0                  0
69200     0.0    mm           0                  0
```

```
                                             thumbnailUrl  \
0           http://www.tate.org.uk/art/images/work/A/A00/A…
1           http://www.tate.org.uk/art/images/work/A/A00/A…
2           http://www.tate.org.uk/art/images/work/A/A00/A…
3           http://www.tate.org.uk/art/images/work/A/A00/A…
4           http://www.tate.org.uk/art/images/work/A/A00/A…
…                                                         …
69196                                                     0
69197                                                     0
69198                                                     0
69199                                                     0
69200       http://www.tate.org.uk/art/images/work/T/T13/T…

                                                         url
0           http://www.tate.org.uk/art/artworks/blake-a-fi…
1           http://www.tate.org.uk/art/artworks/blake-two-…
2           http://www.tate.org.uk/art/artworks/blake-the-…
3           http://www.tate.org.uk/art/artworks/blake-six-…
4           http://www.tate.org.uk/art/artworks/blake-the-…
…                                                         …
69196       http://www.tate.org.uk/art/artworks/p-orridge-…
69197       http://www.tate.org.uk/art/artworks/p-orridge-…
69198       http://www.tate.org.uk/art/artworks/hatoum-pre…
69199       http://www.tate.org.uk/art/artworks/creed-work…
69200       http://www.tate.org.uk/art/artworks/brunias-da…

[69201 rows x 20 columns]
```

```python
# previous data was not modified inplace, so we can still get the original data
# inplace=True can be used to modify original data
depthna = data.fillna(value={'depth': 0})
depth_index = list(data.columns).index('depth')
data.iloc[0:8, depth_index -2: depth_index+2]
```

```
   width  height  depth  units
0    394     419    NaN     mm
1    311     213    NaN     mm
2    343     467    NaN     mm
3    318     394    NaN     mm
4    243     335    NaN     mm
5    240     338    NaN     mm
6    242     334    NaN     mm
7    246     340    NaN     mm
```

```python
depthna.loc[depthna['depth'] == np.nan, ['depth']]
```

```
[ ]: Empty DataFrame
     Columns: [depth]
     Index: []
```

```
[ ]: depthna.loc[depthna['depth'] == 0, ['depth']]
```

```
[ ]:          depth
     0          0.0
     1          0.0
     2          0.0
     3          0.0
     4          0.0
     …           …
     69191      0.0
     69192      0.0
     69193      0.0
     69199      0.0
     69200      0.0

     [66687 rows x 1 columns]
```

1.1 Drop NaN Rows

```
[ ]: data = pd.read_csv('./data/artwork_data.csv')
```

C:\Users\hp\AppData\Local\Temp\ipykernel_13652\74990849.py:1: DtypeWarning:
Columns (9,13) have mixed types. Specify dtype option on import or set
low_memory=False.
  data = pd.read_csv('./data/artwork_data.csv')

```
[ ]: data.shape
```

```
[ ]: (69201, 20)
```

```
[ ]: data.dropna()
     # Equivalent to
     # data.dropna(how='any')
```

```
[ ]:            id accession_number                    artist artistRole  artistId  \
     1738    98171          AR00001             Katz, Alex     artist      1386
     1739    98172          AR00002             Katz, Alex     artist      1386
     1740    98173          AR00003             Katz, Alex     artist      1386
     1741    98174          AR00004             Katz, Alex     artist      1386
     1742    98175          AR00005             Katz, Alex     artist      1386
     …         …              …                     …           …          …
     61878   20155          T06453        Buchheister, Carl    artist      2276
     62083   20568          T06659           Gober, Robert     artist      2309
     62274   20984          T06858  Paolozzi, Sir Eduardo     artist      1738
```

27

```
62466  21517           T07052          Irvin, Albert     artist     1342
62596  21805           T07183          Davey, Grenville  artist     2295

                               title dateText  \
1738                         Pansies     1967
1739                          Tulips     1969
1740                     West Window     1979
1741   Lillies Against Yellow House     1983
1742                     Young Trees     1989
…                                 …       …
61878            Composition Ursiem     1959
62083                         Drains     1990
62274                        Cyclops     1957
62466                     St Germain     1995
62596                         By Air     1989


                                               medium  \
1738                        Oil paint on hardboard
1739                        Oil paint on hardboard
1740                        Oil paint on hardboard
1741                        Oil paint on hardboard
1742                        Oil paint on hardboard
…                                                   …
61878  Graphite, ink, oil paint and adhesive on paper…
62083                                          Pewter
62274                                          Bronze
62466                        Acrylic paint on canvas
62596                     Hardboard, wood and steel


                                           creditLine     year  \
1738   ARTIST ROOMS Acquired jointly with the Nationa…  1967.0
1739   ARTIST ROOMS Acquired jointly with the Nationa…  1969.0
1740   ARTIST ROOMS Acquired jointly with the Nationa…  1979.0
1741   ARTIST ROOMS Acquired jointly with the Nationa…  1983.0
1742   ARTIST ROOMS Acquired jointly with the Nationa…  1989.0
…                                                    …       …
61878                  Presented by Willi Kemp 1991  1959.0
62083                               Purchased 1992  1990.0
62274                               Purchased 1994  1957.0
62466                               Purchased 1996  1995.0
62596      Presented by Janet Wolfson de Botton 1996  1989.0


       acquisitionYear                                    dimensions  \
1738            2008.0  support: 309 x 404 x 4 mm frame: 322 x 424 x 3…
1739            2008.0  support: 356 x 253 x 3 mm frame: 375 x 273 x 3…
1740            2008.0  support: 196 x 238 x 3 mm frame: 216 x 256 x 3…
1741            2008.0  support: 307 x 229 x 3 mm frame: 327 x 249 x 3…
```

```
1742            2008.0  support: 407 x 301 x 3 mm frame: 426 x 320 x 3…
…               …                                                    …
61878           1991.0                        support: 302 x 389 x 23 mm
62083           1992.0                           object: 94 x 94 x 42 mm
62274           1994.0               object: 692 x 370 x 320 mm, 36.4 kg
62466           1996.0                       support: 2140 x 3056 x 52 mm
62596           1996.0                displayed: 1850 x 3250 x 1850 mm


      width height  depth units     inscription  \
1738    309  404.0    4.0    mm  date inscribed
1739    356  253.0    3.0    mm  date inscribed
1740    196  238.0    3.0    mm  date inscribed
1741    307  229.0    3.0    mm  date inscribed
1742    407  301.0    3.0    mm  date inscribed
…       …     …       …      …         …
61878   302   389   23.0    mm  date inscribed
62083    94    94   42.0    mm  date inscribed
62274   692   370  320.0    mm  date inscribed
62466  2140  3056   52.0    mm  date inscribed
62596  1850  3250 1850.0    mm  date inscribed


                     thumbnailCopyright  \
1738                        © Alex Katz
1739                        © Alex Katz
1740                        © Alex Katz
1741                        © Alex Katz
1742                        © Alex Katz
…                                  …
61878                      © DACS, 2014
62083                    © Robert Gober
62274  © The Eduardo Paolozzi Foundation
62466                     © Albert Irvin
62596                  © Grenville Davey


                                    thumbnailUrl  \
1738   http://www.tate.org.uk/art/images/work/AR/AR00…
1739   http://www.tate.org.uk/art/images/work/AR/AR00…
1740   http://www.tate.org.uk/art/images/work/AR/AR00…
1741   http://www.tate.org.uk/art/images/work/AR/AR00…
1742   http://www.tate.org.uk/art/images/work/AR/AR00…
…                                               …
61878  http://www.tate.org.uk/art/images/work/T/T06/T…
62083  http://www.tate.org.uk/art/images/work/T/T06/T…
62274  http://www.tate.org.uk/art/images/work/T/T06/T…
62466  http://www.tate.org.uk/art/images/work/T/T07/T…
62596  http://www.tate.org.uk/art/images/work/T/T07/T…
```

```
                                                          url
1738      http://www.tate.org.uk/art/artworks/katz-pansi…
1739      http://www.tate.org.uk/art/artworks/katz-tulip…
1740      http://www.tate.org.uk/art/artworks/katz-west-…
1741      http://www.tate.org.uk/art/artworks/katz-lilli…
1742      http://www.tate.org.uk/art/artworks/katz-young…
…                                                         …
61878     http://www.tate.org.uk/art/artworks/buchheiste…
62083     http://www.tate.org.uk/art/artworks/gober-drai…
62274     http://www.tate.org.uk/art/artworks/paolozzi-c…
62466     http://www.tate.org.uk/art/artworks/irvin-st-g…
62596     http://www.tate.org.uk/art/artworks/davey-by-a…

[164 rows x 20 columns]
```

[ ]: `data.dropna().shape`

[ ]: (164, 20)

[ ]:
```python
# can be used at the end of processing data
# e.g after setting up na values to different columns, and at the end
# if there is any row which has been set to na totally, just drop that row.
data.dropna(how='all')
```

[ ]:
```
              id accession_number              artist artistRole  artistId  \
0           1035           A00001       Blake, Robert     artist        38
1           1036           A00002       Blake, Robert     artist        38
2           1037           A00003       Blake, Robert     artist        38
3           1038           A00004       Blake, Robert     artist        38
4           1039           A00005      Blake, William     artist        39
…            …                …                   …          …         …
69196     122960           T13865   P-Orridge, Genesis     artist     16646
69197     122961           T13866   P-Orridge, Genesis     artist     16646
69198     121181           T13867        Hatoum, Mona     artist      2365
69199     112306           T13868        Creed, Martin     artist      2760
69200     127035           T13869   Brunias, Agostino     artist      3824

                                                        title  \
0           A Figure Bowing before a Seated Old Man with h…
1           Two Drawings of Frightened Figures, Probably f…
2           The Preaching of Warning. Verso: An Old Man En…
3               Six Drawings of Figures with Outstretched Arms
4           The Circle of the Lustful: Francesca da Rimini…
…                                                           …
69196                         Larvae (from Tampax Romana)
69197                     Living Womb (from Tampax Romana)
69198                                       Present Tense
```

```
69199          Work No. 227: The lights going on and off
69200               Dancing Scene in the West Indies

                          dateText  \
0              date not known
1              date not known
2                      ?c.1785
3              date not known
4       1826-7, reprinted 1892
…                          …
69196                     1975
69197                     1976
69198                     1996
69199                     2000
69200                  1764-96


                                              medium  \
0        Watercolour, ink, chalk and graphite on paper…
1                                      Graphite on paper
2          Graphite on paper. Verso: graphite on paper
3                                      Graphite on paper
4                              Line engraving on paper
…                                                    …
69196    Perspex, Wood, hairpiece, tampon and human blood
69197    Wood, Perspex, plastic, photograph on paper, t…
69198                              Soap and glass beads
69199                                  Gallery lighting
69200                              Oil paint on canvas


                                        creditLine    year  \
0                 Presented by Mrs John Richmond 1922     NaN
1                 Presented by Mrs John Richmond 1922     NaN
2                 Presented by Mrs John Richmond 1922  1785.0
3                 Presented by Mrs John Richmond 1922     NaN
4       Purchased with the assistance of a special gra…  1826.0
…                                                    …     …
69196               Transferred from Tate Archive 2012    1975
69197               Transferred from Tate Archive 2012    1976
69198                   Presented by Tate Members 2013    1996
69199   Purchased with funds provided by Tate Members,…    2000
69200   Purchased with assistance from Tate Patrons an…    1764


     acquisitionYear                      dimensions width  height  \
0            1922.0           support: 394 x 419 mm    394   419.0
1            1922.0           support: 311 x 213 mm    311   213.0
2            1922.0           support: 343 x 467 mm    343   467.0
3            1922.0           support: 318 x 394 mm    318   394.0
```

31

```
4                  1919.0              image: 243 x 335 mm   243    335.0
…                  …                                     …    …     …
69196              2013.0        object: 305 x 305 x 135 mm   305    305.0
69197              2013.0        object: 305 x 305 x 135 mm   305    305.0
69198              2013.0      displayed: 45 x 2410 x 2990 mm   45    2410.0
69199              2013.0  Overall display dimensions variable  NaN    NaN
69200              2013.0         unconfirmed: 508 x 660 mm   508    660.0


        depth units inscription thumbnailCopyright  \
0          NaN    mm        NaN                NaN
1          NaN    mm        NaN                NaN
2          NaN    mm        NaN                NaN
3          NaN    mm        NaN                NaN
4          NaN    mm        NaN                NaN

…          …     …         …                  …
69196    135.0    mm        NaN                NaN
69197    135.0    mm        NaN                NaN
69198   2990.0    mm        NaN                NaN
69199      NaN   NaN        NaN                NaN
69200      NaN    mm        NaN                NaN


                                      thumbnailUrl  \
0      http://www.tate.org.uk/art/images/work/A/A00/A…
1      http://www.tate.org.uk/art/images/work/A/A00/A…
2      http://www.tate.org.uk/art/images/work/A/A00/A…
3      http://www.tate.org.uk/art/images/work/A/A00/A…
4      http://www.tate.org.uk/art/images/work/A/A00/A…
…                                               …
69196                                           NaN
69197                                           NaN
69198                                           NaN
69199                                           NaN
69200  http://www.tate.org.uk/art/images/work/T/T13/T…


                                      url
0      http://www.tate.org.uk/art/artworks/blake-a-fi…
1      http://www.tate.org.uk/art/artworks/blake-two-…
2      http://www.tate.org.uk/art/artworks/blake-the-…
3      http://www.tate.org.uk/art/artworks/blake-six-…
4      http://www.tate.org.uk/art/artworks/blake-the-…
…                                               …
69196  http://www.tate.org.uk/art/artworks/p-orridge-…
69197  http://www.tate.org.uk/art/artworks/p-orridge-…
69198  http://www.tate.org.uk/art/artworks/hatoum-pre…
69199  http://www.tate.org.uk/art/artworks/creed-work…
69200  http://www.tate.org.uk/art/artworks/brunias-da…
```

```
[69201 rows x 20 columns]
```

```python
print(data.shape)
print(data.dropna(how='all').shape)
```

```
(69201, 20)
(69201, 20)
```

```python
len(list(data.columns))
```

```
20
```

```python
# drop rows using a thresh value

# thresh : int, optional
# Require that many non-NA values.
data.dropna(thresh=15) # at least 15 values should not be nan values
```

```
            id accession_number              artist artistRole  artistId  \
0         1035           A00001      Blake, Robert      artist        38
1         1036           A00002      Blake, Robert      artist        38
2         1037           A00003      Blake, Robert      artist        38
3         1038           A00004      Blake, Robert      artist        38
4         1039           A00005      Blake, William     artist        39
...        ...              ...                ...        ...       ...
69195   122959           T13864  P-Orridge, Genesis     artist     16646
69196   122960           T13865  P-Orridge, Genesis     artist     16646
69197   122961           T13866  P-Orridge, Genesis     artist     16646
69198   121181           T13867       Hatoum, Mona     artist      2365
69200   127035           T13869  Brunias, Agostino     artist      3824

                                                   title  \
0          A Figure Bowing before a Seated Old Man with h…
1          Two Drawings of Frightened Figures, Probably f…
2          The Preaching of Warning. Verso: An Old Man En…
3            Six Drawings of Figures with Outstretched Arms
4          The Circle of the Lustful: Francesca da Rimini…
...                                                    …
69195     It's That Time Of The Month (from Tampax Romana)
69196                   Larvae (from Tampax Romana)
69197              Living Womb (from Tampax Romana)
69198                                    Present Tense
69200                   Dancing Scene in the West Indies

                  dateText  \
0            date not known
1            date not known
2                   ?c.1785
```

```
3                         date not known
4         1826-7, reprinted 1892
…                             …
69195                          1975
69196                          1975
69197                          1976
69198                          1996
69200                       1764-96


                                                              medium  \
0        Watercolour, ink, chalk and graphite on paper…
1                                    Graphite on paper
2           Graphite on paper. Verso: graphite on paper
3                                    Graphite on paper
4                            Line engraving on paper
…                                       …
69195   Wood, Perspex, clock case, tampons and human b…
69196    Perspex, Wood, hairpiece, tampon and human blood
69197   Wood, Perspex, plastic, photograph on paper, t…
69198                           Soap and glass beads
69200                            Oil paint on canvas


                                           creditLine    year  \
0                   Presented by Mrs John Richmond 1922     NaN
1                   Presented by Mrs John Richmond 1922     NaN
2                   Presented by Mrs John Richmond 1922  1785.0
3                   Presented by Mrs John Richmond 1922     NaN
4         Purchased with the assistance of a special gra…  1826.0
…                                          …            …
69195               Transferred from Tate Archive 2012     1975
69196               Transferred from Tate Archive 2012     1975
69197               Transferred from Tate Archive 2012     1976
69198                 Presented by Tate Members 2013      1996
69200   Purchased with assistance from Tate Patrons an…     1764


        acquisitionYear                    dimensions width   height   depth  \
0               1922.0         support: 394 x 419 mm    394    419.0     NaN
1               1922.0         support: 311 x 213 mm    311    213.0     NaN
2               1922.0         support: 343 x 467 mm    343    467.0     NaN
3               1922.0         support: 318 x 394 mm    318    394.0     NaN
4               1919.0           image: 243 x 335 mm    243    335.0     NaN
…                  …                       …          …       …         …
69195           2013.0     object: 305 x 305 x 135 mm   305    305.0    135.0
69196           2013.0     object: 305 x 305 x 135 mm   305    305.0    135.0
69197           2013.0     object: 305 x 305 x 135 mm   305    305.0    135.0
69198           2013.0  displayed: 45 x 2410 x 2990 mm    45   2410.0   2990.0
69200           2013.0      unconfirmed: 508 x 660 mm    508    660.0     NaN
```

```
       units inscription   thumbnailCopyright  \
0         mm         NaN                   NaN
1         mm         NaN                   NaN
2         mm         NaN                   NaN
3         mm         NaN                   NaN
4         mm         NaN                   NaN
...       ...        ...                   ...
69195     mm         NaN  © Genesis P-Orridge
69196     mm         NaN                   NaN
69197     mm         NaN                   NaN
69198     mm         NaN                   NaN
69200     mm         NaN                   NaN

                                           thumbnailUrl  \
0         http://www.tate.org.uk/art/images/work/A/A00/A…
1         http://www.tate.org.uk/art/images/work/A/A00/A…
2         http://www.tate.org.uk/art/images/work/A/A00/A…
3         http://www.tate.org.uk/art/images/work/A/A00/A…
4         http://www.tate.org.uk/art/images/work/A/A00/A…
...                                                    …
69195     http://www.tate.org.uk/art/images/work/T/T13/T…
69196                                                NaN
69197                                                NaN
69198                                                NaN
69200     http://www.tate.org.uk/art/images/work/T/T13/T…

                                                    url
0         http://www.tate.org.uk/art/artworks/blake-a-fi…
1         http://www.tate.org.uk/art/artworks/blake-two-…
2         http://www.tate.org.uk/art/artworks/blake-the-…
3         http://www.tate.org.uk/art/artworks/blake-six-…
4         http://www.tate.org.uk/art/artworks/blake-the-…
...                                                    …
69195     http://www.tate.org.uk/art/artworks/p-orridge-…
69196     http://www.tate.org.uk/art/artworks/p-orridge-…
69197     http://www.tate.org.uk/art/artworks/p-orridge-…
69198     http://www.tate.org.uk/art/artworks/hatoum-pre…
69200     http://www.tate.org.uk/art/artworks/brunias-da…

[66379 rows x 20 columns]
```

```python
help(data.dropna)
```

```
Help on method dropna in module pandas.core.frame:

dropna(axis: 'Axis' = 0, how: 'str' = 'any', thresh=None, subset: 'IndexLabel' =
None, inplace: 'bool' = False) method of pandas.core.frame.DataFrame instance
```

Remove missing values.

See the :ref:`User Guide <missing_data>` for more on which values are
considered missing, and how to work with missing data.

Parameters
----
axis : {0 or 'index', 1 or 'columns'}, default 0
    Determine if rows or columns which contain missing values are
    removed.

    * 0, or 'index' : Drop rows which contain missing values.
    * 1, or 'columns' : Drop columns which contain missing value.

    .. versionchanged:: 1.0.0

       Pass tuple or list to drop on multiple axes.
       Only a single axis is allowed.

how : {'any', 'all'}, default 'any'
    Determine if row or column is removed from DataFrame, when we have
    at least one NA or all NA.

    * 'any' : If any NA values are present, drop that row or column.
    * 'all' : If all values are NA, drop that row or column.

thresh : int, optional
    Require that many non-NA values.
subset : column label or sequence of labels, optional
    Labels along other axis to consider, e.g. if you are dropping rows
    these would be a list of columns to include.
inplace : bool, default False
    If True, do operation inplace and return None.

Returns

DataFrame or None
    DataFrame with NA entries dropped from it or None if ``inplace=True``.

See Also

DataFrame.isna: Indicate missing values.
DataFrame.notna : Indicate existing (non-missing) values.
DataFrame.fillna : Replace missing values.
Series.dropna : Drop missing values.
Index.dropna : Drop missing indices.

Examples

```
--------
>>> df = pd.DataFrame({"name": ['Alfred', 'Batman', 'Catwoman'],
…                      "toy": [np.nan, 'Batmobile', 'Bullwhip'],
…                      "born": [pd.NaT, pd.Timestamp("1940-04-25"),
…                               pd.NaT]})
>>> df
       name        toy       born
0    Alfred        NaN        NaT
1    Batman  Batmobile 1940-04-25
2  Catwoman   Bullwhip        NaT

Drop the rows where at least one element is missing.

>>> df.dropna()
     name        toy       born
1  Batman  Batmobile 1940-04-25

Drop the columns where at least one element is missing.

>>> df.dropna(axis='columns')
       name
0    Alfred
1    Batman
2  Catwoman

Drop the rows where all elements are missing.

>>> df.dropna(how='all')
       name        toy       born
0    Alfred        NaN        NaT
1    Batman  Batmobile 1940-04-25
2  Catwoman   Bullwhip        NaT

Keep only the rows with at least 2 non-NA values.

>>> df.dropna(thresh=2)
       name        toy       born
1    Batman  Batmobile 1940-04-25
2  Catwoman   Bullwhip        NaT

Define in which columns to look for missing values.

>>> df.dropna(subset=['name', 'toy'])
       name        toy       born
1    Batman  Batmobile 1940-04-25
2  Catwoman   Bullwhip        NaT

Keep the DataFrame with valid entries in the same variable.
```

```
>>> df.dropna(inplace=True)
>>> df
       name        toy        born
1    Batman  Batmobile  1940-04-25
```

```
[ ]: # inplace=True can be passed to drop in place
     data.dropna(subset=['year', 'acquisitionYear'], how='all')
```

```
[ ]:              id accession_number               artist artistRole  artistId  \
     0          1035           A00001        Blake, Robert     artist        38
     1          1036           A00002        Blake, Robert     artist        38
     2          1037           A00003        Blake, Robert     artist        38
     3          1038           A00004        Blake, Robert     artist        38
     4          1039           A00005       Blake, William     artist        39
     ...         ...              ...                  ...        ...       ...
     69196    122960           T13865    P-Orridge, Genesis     artist     16646
     69197    122961           T13866    P-Orridge, Genesis     artist     16646
     69198    121181           T13867          Hatoum, Mona     artist      2365
     69199    112306           T13868         Creed, Martin     artist      2760
     69200    127035           T13869     Brunias, Agostino     artist      3824

                                                        title  \
     0          A Figure Bowing before a Seated Old Man with h…
     1          Two Drawings of Frightened Figures, Probably f…
     2          The Preaching of Warning. Verso: An Old Man En…
     3              Six Drawings of Figures with Outstretched Arms
     4          The Circle of the Lustful: Francesca da Rimini…
     ...                                                     …
     69196                        Larvae (from Tampax Romana)
     69197                   Living Womb (from Tampax Romana)
     69198                                      Present Tense
     69199            Work No. 227: The lights going on and off
     69200                      Dancing Scene in the West Indies

                         dateText  \
     0            date not known
     1            date not known
     2                   ?c.1785
     3            date not known
     4       1826-7, reprinted 1892
     ...                        …
     69196                  1975
     69197                  1976
     69198                  1996
     69199                  2000
```

```
69200                     1764-96


                                                          medium  \
0       Watercolour, ink, chalk and graphite on paper…
1                                    Graphite on paper
2           Graphite on paper. Verso: graphite on paper
3                                    Graphite on paper
4                             Line engraving on paper
…                                                    …
69196  Perspex, Wood, hairpiece, tampon and human blood
69197  Wood, Perspex, plastic, photograph on paper, t…
69198                              Soap and glass beads
69199                                  Gallery lighting
69200                              Oil paint on canvas


                                          creditLine    year  \
0               Presented by Mrs John Richmond 1922     NaN
1               Presented by Mrs John Richmond 1922     NaN
2               Presented by Mrs John Richmond 1922  1785.0
3               Presented by Mrs John Richmond 1922     NaN
4       Purchased with the assistance of a special gra…  1826.0
…                                                    …     …
69196             Transferred from Tate Archive 2012    1975
69197             Transferred from Tate Archive 2012    1976
69198              Presented by Tate Members 2013      1996
69199  Purchased with funds provided by Tate Members,…    2000
69200  Purchased with assistance from Tate Patrons an…    1764


       acquisitionYear                        dimensions width   height  \
0              1922.0              support: 394 x 419 mm    394   419.0
1              1922.0              support: 311 x 213 mm    311   213.0
2              1922.0              support: 343 x 467 mm    343   467.0
3              1922.0              support: 318 x 394 mm    318   394.0
4              1919.0                image: 243 x 335 mm    243   335.0
…                  …                                 …     …       …
69196          2013.0          object: 305 x 305 x 135 mm  305   305.0
69197          2013.0          object: 305 x 305 x 135 mm  305   305.0
69198          2013.0       displayed: 45 x 2410 x 2990 mm  45  2410.0
69199          2013.0  Overall display dimensions variable  NaN     NaN
69200          2013.0            unconfirmed: 508 x 660 mm  508   660.0


       depth units inscription thumbnailCopyright  \
0       NaN    mm         NaN                NaN
1       NaN    mm         NaN                NaN
2       NaN    mm         NaN                NaN
3       NaN    mm         NaN                NaN
4       NaN    mm         NaN                NaN
```

```
…       …    …         …                   …
69196   135.0    mm         NaN                 NaN
69197   135.0    mm         NaN                 NaN
69198  2990.0    mm         NaN                 NaN
69199     NaN   NaN         NaN                 NaN
69200     NaN    mm         NaN                 NaN


                                                thumbnailUrl  \
0        http://www.tate.org.uk/art/images/work/A/A00/A…
1        http://www.tate.org.uk/art/images/work/A/A00/A…
2        http://www.tate.org.uk/art/images/work/A/A00/A…
3        http://www.tate.org.uk/art/images/work/A/A00/A…
4        http://www.tate.org.uk/art/images/work/A/A00/A…
…                                                       …
69196                                                 NaN
69197                                                 NaN
69198                                                 NaN
69199                                                 NaN
69200  http://www.tate.org.uk/art/images/work/T/T13/T…


                                                         url
0        http://www.tate.org.uk/art/artworks/blake-a-fi…
1        http://www.tate.org.uk/art/artworks/blake-two-…
2        http://www.tate.org.uk/art/artworks/blake-the-…
3        http://www.tate.org.uk/art/artworks/blake-six-…
4        http://www.tate.org.uk/art/artworks/blake-the-…
…                                                         …
69196  http://www.tate.org.uk/art/artworks/p-orridge-…
69197  http://www.tate.org.uk/art/artworks/p-orridge-…
69198  http://www.tate.org.uk/art/artworks/hatoum-pre…
69199  http://www.tate.org.uk/art/artworks/creed-work…
69200  http://www.tate.org.uk/art/artworks/brunias-da…

[69198 rows x 20 columns]
```

1.2   Identifying and Drop Duplicates

```
[ ]: data = pd.read_csv('./data/artwork_sample.csv')
```

```
[ ]: data
```

```
[ ]:      id accession_number          artist artistRole  artistId  \
     0  1035           A00001   Blake, Robert     artist        38
     1  1036           A00002   Blake, Robert     artist        38
     2  1037           A00003   Blake, Robert     artist        38
     3  1038           A00004   Blake, Robert     artist        38
     4  1039           A00005  Blake, William     artist        39
```

```
5  1040          A00006  Blake, William    artist         39
6  1041          A00007  Blake, William    artist         39
7  1042          A00008  Blake, William    artist         39
8  1043          A00009  Blake, William    artist         39
9  1044          A00010  Blake, William    artist         39

                                          title              dateText  \
0  A Figure Bowing before a Seated Old Man with h…      date not known
1  Two Drawings of Frightened Figures, Probably f…      date not known
2  The Preaching of Warning. Verso: An Old Man En…              ?c.1785
3    Six Drawings of Figures with Outstretched Arms      date not known
4  The Circle of the Lustful: Francesca da Rimini…  1826-7, reprinted 1892
5      Ciampolo the Barrator Tormented by the Devils  1826-7, reprinted 1892
6                      The Baffled Devils Fighting  1826-7, reprinted 1892
7  The Six-Footed Serpent Attacking Agnolo Brunel…  1826-7, reprinted 1892
8              The Serpent Attacking Buoso Donati  1826-7, reprinted 1892
9                The Pit of Disease: The Falsifiers  1826-7, reprinted 1892

                                          medium  \
0  Watercolour, ink, chalk and graphite on paper…
1                              Graphite on paper
2        Graphite on paper. Verso: graphite on paper
3                              Graphite on paper
4                        Line engraving on paper
5                        Line engraving on paper
6                        Line engraving on paper
7                        Line engraving on paper
8                        Line engraving on paper
9                        Line engraving on paper

                                      creditLine    year  acquisitionYear  \
0              Presented by Mrs John Richmond 1922     NaN             1922
1              Presented by Mrs John Richmond 1922     NaN             1922
2              Presented by Mrs John Richmond 1922  1785.0             1922
3              Presented by Mrs John Richmond 1922     NaN             1922
4  Purchased with the assistance of a special gra…  1826.0             1919
5  Purchased with the assistance of a special gra…  1826.0             1919
6  Purchased with the assistance of a special gra…  1826.0             1919
7  Purchased with the assistance of a special gra…  1826.0             1919
8  Purchased with the assistance of a special gra…  1826.0             1919
9  Purchased with the assistance of a special gra…  1826.0             1919

              dimensions  width  height  depth units  inscription  \
0  support: 394 x 419 mm  394.0   419.0    NaN    mm          NaN
1  support: 311 x 213 mm  311.0   213.0    NaN    mm          NaN
2  support: 343 x 467 mm  343.0   467.0    NaN    mm          NaN
3  support: 318 x 394 mm  318.0   394.0    NaN    mm          NaN
```

```
4     image: 243 x 335 mm   243.0   335.0   NaN   mm        NaN
5     image: 240 x 338 mm   240.0   338.0   NaN   mm        NaN
6     image: 242 x 334 mm   242.0   334.0   NaN   mm        NaN
7     image: 246 x 340 mm   246.0   340.0   NaN   mm        NaN
8     image: 241 x 335 mm   241.0   335.0   NaN   mm        NaN
9     image: 243 x 340 mm   243.0   340.0   NaN   mm        NaN

    thumbnailCopyright                             thumbnailUrl  \
0               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
1               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
2               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
3               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
4               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
5               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
6               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
7               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
8               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
9               NaN  http://www.tate.org.uk/art/images/work/A/A00/A…

                                                url
0  http://www.tate.org.uk/art/artworks/blake-a-fi…
1  http://www.tate.org.uk/art/artworks/blake-two-…
2  http://www.tate.org.uk/art/artworks/blake-the-…
3  http://www.tate.org.uk/art/artworks/blake-six-…
4  http://www.tate.org.uk/art/artworks/blake-the-…
5  http://www.tate.org.uk/art/artworks/blake-ciam…
6  http://www.tate.org.uk/art/artworks/blake-the-…
7  http://www.tate.org.uk/art/artworks/blake-the-…
8  http://www.tate.org.uk/art/artworks/blake-the-…
9  http://www.tate.org.uk/art/artworks/blake-the-…
```

```
[ ]: data.drop_duplicates()
```

```
[ ]:      id accession_number          artist artistRole  artistId  \
     0  1035           A00001   Blake, Robert     artist        38
     1  1036           A00002   Blake, Robert     artist        38
     2  1037           A00003   Blake, Robert     artist        38
     3  1038           A00004   Blake, Robert     artist        38
     4  1039           A00005  Blake, William     artist        39
     5  1040           A00006  Blake, William     artist        39
     6  1041           A00007  Blake, William     artist        39
     7  1042           A00008  Blake, William     artist        39
     8  1043           A00009  Blake, William     artist        39
     9  1044           A00010  Blake, William     artist        39

                                                title          dateText  \
     0  A Figure Bowing before a Seated Old Man with h…     date not known
```

```
1  Two Drawings of Frightened Figures, Probably f…          date not known
2  The Preaching of Warning. Verso: An Old Man En…                 ?c.1785
3      Six Drawings of Figures with Outstretched Arms          date not known
4  The Circle of the Lustful: Francesca da Rimini…  1826-7, reprinted 1892
5      Ciampolo the Barrator Tormented by the Devils  1826-7, reprinted 1892
6                         The Baffled Devils Fighting  1826-7, reprinted 1892
7  The Six-Footed Serpent Attacking Agnolo Brunel…  1826-7, reprinted 1892
8                  The Serpent Attacking Buoso Donati  1826-7, reprinted 1892
9                  The Pit of Disease: The Falsifiers  1826-7, reprinted 1892


                                                 medium  \
0  Watercolour, ink, chalk and graphite on paper…
1                                     Graphite on paper
2         Graphite on paper. Verso: graphite on paper
3                                     Graphite on paper
4                               Line engraving on paper
5                               Line engraving on paper
6                               Line engraving on paper
7                               Line engraving on paper
8                               Line engraving on paper
9                               Line engraving on paper


                                          creditLine    year  acquisitionYear  \
0              Presented by Mrs John Richmond 1922     NaN             1922
1              Presented by Mrs John Richmond 1922     NaN             1922
2              Presented by Mrs John Richmond 1922  1785.0             1922
3              Presented by Mrs John Richmond 1922     NaN             1922
4  Purchased with the assistance of a special gra…  1826.0             1919
5  Purchased with the assistance of a special gra…  1826.0             1919
6  Purchased with the assistance of a special gra…  1826.0             1919
7  Purchased with the assistance of a special gra…  1826.0             1919
8  Purchased with the assistance of a special gra…  1826.0             1919
9  Purchased with the assistance of a special gra…  1826.0             1919


              dimensions  width  height  depth units  inscription  \
0  support: 394 x 419 mm  394.0   419.0    NaN    mm          NaN
1  support: 311 x 213 mm  311.0   213.0    NaN    mm          NaN
2  support: 343 x 467 mm  343.0   467.0    NaN    mm          NaN
3  support: 318 x 394 mm  318.0   394.0    NaN    mm          NaN
4    image: 243 x 335 mm  243.0   335.0    NaN    mm          NaN
5    image: 240 x 338 mm  240.0   338.0    NaN    mm          NaN
6    image: 242 x 334 mm  242.0   334.0    NaN    mm          NaN
7    image: 246 x 340 mm  246.0   340.0    NaN    mm          NaN
8    image: 241 x 335 mm  241.0   335.0    NaN    mm          NaN
9    image: 243 x 340 mm  243.0   340.0    NaN    mm          NaN


  thumbnailCopyright                                       thumbnailUrl  \
```

```
0              NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
1              NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
2              NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
3              NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
4              NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
5              NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
6              NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
7              NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
8              NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
9              NaN  http://www.tate.org.uk/art/images/work/A/A00/A…

                                             url
0  http://www.tate.org.uk/art/artworks/blake-a-fi…
1  http://www.tate.org.uk/art/artworks/blake-two-…
2  http://www.tate.org.uk/art/artworks/blake-the-…
3  http://www.tate.org.uk/art/artworks/blake-six-…
4  http://www.tate.org.uk/art/artworks/blake-the-…
5  http://www.tate.org.uk/art/artworks/blake-ciam…
6  http://www.tate.org.uk/art/artworks/blake-the-…
7  http://www.tate.org.uk/art/artworks/blake-the-…
8  http://www.tate.org.uk/art/artworks/blake-the-…
9  http://www.tate.org.uk/art/artworks/blake-the-…
```

```
[ ]: data.drop_duplicates(subset=['artist'])
```

```
[ ]:       id accession_number           artist artistRole  artistId  \
     0  1035           A00001    Blake, Robert     artist        38
     4  1039           A00005  Blake, William     artist        39

                                         title             dateText  \
     0  A Figure Bowing before a Seated Old Man with h…      date not known
     4  The Circle of the Lustful: Francesca da Rimini…  1826-7, reprinted 1892

                                      medium  \
     0  Watercolour, ink, chalk and graphite on paper…
     4                    Line engraving on paper

                                    creditLine     year  acquisitionYear  \
     0            Presented by Mrs John Richmond 1922      NaN             1922
     4  Purchased with the assistance of a special gra…  1826.0             1919

                 dimensions  width  height  depth units  inscription  \
     0  support: 394 x 419 mm  394.0   419.0    NaN    mm          NaN
     4    image: 243 x 335 mm  243.0   335.0    NaN    mm          NaN

        thumbnailCopyright                                    thumbnailUrl  \
     0              NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
```

```
4                    NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
```

```
                                                 url
0  http://www.tate.org.uk/art/artworks/blake-a-fi…
4  http://www.tate.org.uk/art/artworks/blake-the-…
```

`[ ]:` `data.drop_duplicates(subset=['artist'], keep='first')`

`[ ]:`
```
     id accession_number          artist artistRole  artistId  \
0  1035           A00001   Blake, Robert     artist        38
4  1039           A00005  Blake, William     artist        39
```

```
                                               title               dateText  \
0  A Figure Bowing before a Seated Old Man with h…          date not known
4  The Circle of the Lustful: Francesca da Rimini…  1826-7, reprinted 1892
```

```
                                       medium  \
0  Watercolour, ink, chalk and graphite on paper…
4                        Line engraving on paper
```

```
                                       creditLine    year  acquisitionYear  \
0              Presented by Mrs John Richmond 1922     NaN             1922
4  Purchased with the assistance of a special gra…  1826.0             1919
```

```
             dimensions  width  height  depth units  inscription  \
0  support: 394 x 419 mm  394.0   419.0    NaN    mm          NaN
4    image: 243 x 335 mm  243.0   335.0    NaN    mm          NaN
```

```
   thumbnailCopyright                                       thumbnailUrl  \
0                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
4                 NaN  http://www.tate.org.uk/art/images/work/A/A00/A…
```

```
                                                 url
0  http://www.tate.org.uk/art/artworks/blake-a-fi…
4  http://www.tate.org.uk/art/artworks/blake-the-…
```

`[ ]:` `data.drop_duplicates(subset=['artist'], keep='last')`

`[ ]:`
```
     id accession_number          artist artistRole  artistId  \
3  1038           A00004   Blake, Robert     artist        38
9  1044           A00010  Blake, William     artist        39
```

```
                                            title                dateText  \
3  Six Drawings of Figures with Outstretched Arms          date not known
9             The Pit of Disease: The Falsifiers  1826-7, reprinted 1892
```

```
                            medium                                     creditLine  \
```

```
3        Graphite on paper                    Presented by Mrs John Richmond 1922
9  Line engraving on paper  Purchased with the assistance of a special gra…

     year  acquisitionYear           dimensions  width  height  depth units  \
3     NaN             1922  support: 318 x 394 mm  318.0   394.0    NaN    mm
9  1826.0             1919    image: 243 x 340 mm  243.0   340.0    NaN    mm


   inscription  thumbnailCopyright  \
3         NaN                 NaN
9         NaN                 NaN


                                       thumbnailUrl  \
3  http://www.tate.org.uk/art/images/work/A/A00/A…
9  http://www.tate.org.uk/art/images/work/A/A00/A…


                                       url
3  http://www.tate.org.uk/art/artworks/blake-six-…
9  http://www.tate.org.uk/art/artworks/blake-the-…
```

```python
# keep : {'first', 'last', False}, default 'first'
# Determines which duplicates (if any) to keep.
# - ``first`` : Drop duplicates except for the first occurrence.
# - ``last`` : Drop duplicates except for the last occurrence.
# - False : Drop all duplicates.

# if duplicate values found based on artist, dont keep any of the rows
data.drop_duplicates(subset=['artist'], keep=False)
```

```
Empty DataFrame
Columns: [id, accession_number, artist, artistRole, artistId, title, dateText,
medium, creditLine, year, acquisitionYear, dimensions, width, height, depth,
units, inscription, thumbnailCopyright, thumbnailUrl, url]
Index: []
```

```python
help(data.drop_duplicates)
```

```
Help on method drop_duplicates in module pandas.core.frame:

drop_duplicates(subset: 'Hashable | Sequence[Hashable] | None' = None, keep:
"Literal['first'] | Literal['last'] | Literal[False]" = 'first', inplace: 'bool'
= False, ignore_index: 'bool' = False) -> 'DataFrame | None' method of
pandas.core.frame.DataFrame instance
    Return DataFrame with duplicate rows removed.

    Considering certain columns is optional. Indexes, including time indexes
    are ignored.

    Parameters
```

```
----------
subset : column label or sequence of labels, optional
    Only consider certain columns for identifying duplicates, by
    default use all of the columns.
keep : {'first', 'last', False}, default 'first'
    Determines which duplicates (if any) to keep.
    - ``first`` : Drop duplicates except for the first occurrence.
    - ``last`` : Drop duplicates except for the last occurrence.
    - False : Drop all duplicates.
inplace : bool, default False
    Whether to drop duplicates in place or to return a copy.
ignore_index : bool, default False
    If True, the resulting axis will be labeled 0, 1, …, n - 1.

    .. versionadded:: 1.0.0


Returns
-------
DataFrame or None
    DataFrame with duplicates removed or None if ``inplace=True``.

See Also
--------
DataFrame.value_counts: Count unique combinations of columns.

Examples
--------
Consider dataset containing ramen rating.

>>> df = pd.DataFrame({
…      'brand': ['Yum Yum', 'Yum Yum', 'Indomie', 'Indomie', 'Indomie'],
…      'style': ['cup', 'cup', 'cup', 'pack', 'pack'],
…      'rating': [4, 4, 3.5, 15, 5]
… })
>>> df
     brand style  rating
0  Yum Yum   cup     4.0
1  Yum Yum   cup     4.0
2  Indomie   cup     3.5
3  Indomie  pack    15.0
4  Indomie  pack     5.0

By default, it removes duplicate rows based on all columns.

>>> df.drop_duplicates()
     brand style  rating
0  Yum Yum   cup     4.0
2  Indomie   cup     3.5
```

```
3  Indomie  pack    15.0
4  Indomie  pack     5.0

To remove duplicates on specific column(s), use ``subset``.

>>> df.drop_duplicates(subset=['brand'])
    brand style  rating
0  Yum Yum   cup     4.0
2  Indomie   cup     3.5

To remove duplicates and keep last occurrences, use ``keep``.

>>> df.drop_duplicates(subset=['brand', 'style'], keep='last')
    brand style  rating
1  Yum Yum   cup     4.0
2  Indomie   cup     3.5
4  Indomie  pack     5.0
```

[ ]: `data = pd.read_csv('./data/artwork_data.csv')`

```
C:\Users\hp\AppData\Local\Temp\ipykernel_13652\74990849.py:1: DtypeWarning:
Columns (9,13) have mixed types. Specify dtype option on import or set
low_memory=False.
  data = pd.read_csv('./data/artwork_data.csv')
```

[ ]: `data.shape`

[ ]: (69201, 20)

[ ]: `data.duplicated()`

```
[ ]: 0        False
     1        False
     2        False
     3        False
     4        False
              …
     69196    False
     69197    False
     69198    False
     69199    False
     69200    False
     Length: 69201, dtype: bool
```

[ ]: `data.loc[data.duplicated()]`
```

```
[ ]: Empty DataFrame
     Columns: [id, accession_number, artist, artistRole, artistId, title, dateText,
     medium, creditLine, year, acquisitionYear, dimensions, width, height, depth,
     units, inscription, thumbnailCopyright, thumbnailUrl, url]
     Index: []
```

```
[ ]: data.duplicated(subset=['artist', 'title'], keep=False)
```

```
[ ]: 0        False
     1        False
     2        False
     3        False
     4         True
              …
     69196    False
     69197    False
     69198    False
     69199    False
     69200    False
     Length: 69201, dtype: bool
```

```
[ ]: data.loc[data.duplicated(subset=['artist', 'title'], keep=False)]
```

```
[ ]:            id accession_number           artist artistRole  artistId  \
     4        1039           A00005   Blake, William     artist        39
     5        1040           A00006   Blake, William     artist        39
     6        1041           A00007   Blake, William     artist        39
     7        1042           A00008   Blake, William     artist        39
     8        1043           A00009   Blake, William     artist        39
     …         …                …               …          …         …
     69169  121194           T13838  Barlow, Phyllida     artist     10908
     69170  121195           T13839  Barlow, Phyllida     artist     10908
     69171  121196           T13840  Barlow, Phyllida     artist     10908
     69172  121197           T13841  Barlow, Phyllida     artist     10908
     69173  121198           T13842  Barlow, Phyllida     artist     10908

                                                          title  \
     4          The Circle of the Lustful: Francesca da Rimini…
     5              Ciampolo the Barrator Tormented by the Devils
     6                            The Baffled Devils Fighting
     7          The Six-Footed Serpent Attacking Agnolo Brunel…
     8                       The Serpent Attacking Buoso Donati
     …                                                        …
     69169                                            Untitled
     69170                                            Untitled
     69171                                            Untitled
     69172                                            Untitled
```

```
69173                                   Untitled

                    dateText                    medium  \
4       1826-7, reprinted 1892  Line engraving on paper
5       1826-7, reprinted 1892  Line engraving on paper
6       1826-7, reprinted 1892  Line engraving on paper
7       1826-7, reprinted 1892  Line engraving on paper
8       1826-7, reprinted 1892  Line engraving on paper
…                           …                        …
69169                     1999    Acrylic paint on paper
69170                     2003    Acrylic paint on paper
69171                     2003    Acrylic paint on paper
69172                     2004    Acrylic paint on paper
69173                     2006    Acrylic paint on paper


                                        creditLine     year  \
4       Purchased with the assistance of a special gra…  1826.0
5       Purchased with the assistance of a special gra…  1826.0
6       Purchased with the assistance of a special gra…  1826.0
7       Purchased with the assistance of a special gra…  1826.0
8       Purchased with the assistance of a special gra…  1826.0
…                                               …        …
69169   Presented by the Trustees of the Chantrey Bequ…    1999
69170   Presented by the Trustees of the Chantrey Bequ…    2003
69171   Presented by the Trustees of the Chantrey Bequ…    2003
69172   Presented by the Trustees of the Chantrey Bequ…    2004
69173   Presented by the Trustees of the Chantrey Bequ…    2006


        acquisitionYear                  dimensions  width  height  depth units  \
4               1919.0        image: 243 x 335 mm     243   335.0    NaN    mm
5               1919.0        image: 240 x 338 mm     240   338.0    NaN    mm
6               1919.0        image: 242 x 334 mm     242   334.0    NaN    mm
7               1919.0        image: 246 x 340 mm     246   340.0    NaN    mm
8               1919.0        image: 241 x 335 mm     241   335.0    NaN    mm
…                   …                          …      …       …      …     …
69169           2013.0  unconfirmed: 637 x 897 mm     637   897.0    NaN    mm
69170           2013.0     support: 560 x 762 mm     560   762.0    NaN    mm
69171           2013.0     support: 558 x 760 mm     558   760.0    NaN    mm
69172           2013.0     support: 380 x 566 mm     380   566.0    NaN    mm
69173           2013.0     support: 560 x 762 mm     560   762.0    NaN    mm


        inscription thumbnailCopyright  \
4               NaN                NaN
5               NaN                NaN
6               NaN                NaN
7               NaN                NaN
8               NaN                NaN
```

```
          …              …                  …
69169            NaN   © Phyllida Barlow
69170            NaN   © Phyllida Barlow
69171            NaN   © Phyllida Barlow
69172            NaN   © Phyllida Barlow
69173            NaN   © Phyllida Barlow


                                        thumbnailUrl  \
4      http://www.tate.org.uk/art/images/work/A/A00/A…
5      http://www.tate.org.uk/art/images/work/A/A00/A…
6      http://www.tate.org.uk/art/images/work/A/A00/A…
7      http://www.tate.org.uk/art/images/work/A/A00/A…
8      http://www.tate.org.uk/art/images/work/A/A00/A…
…                                                   …
69169  http://www.tate.org.uk/art/images/work/T/T13/T…
69170  http://www.tate.org.uk/art/images/work/T/T13/T…
69171  http://www.tate.org.uk/art/images/work/T/T13/T…
69172  http://www.tate.org.uk/art/images/work/T/T13/T…
69173  http://www.tate.org.uk/art/images/work/T/T13/T…


                                                 url
4      http://www.tate.org.uk/art/artworks/blake-the-…
5      http://www.tate.org.uk/art/artworks/blake-ciam…
6      http://www.tate.org.uk/art/artworks/blake-the-…
7      http://www.tate.org.uk/art/artworks/blake-the-…
8      http://www.tate.org.uk/art/artworks/blake-the-…
…                                                   …
69169  http://www.tate.org.uk/art/artworks/barlow-unt…
69170  http://www.tate.org.uk/art/artworks/barlow-unt…
69171  http://www.tate.org.uk/art/artworks/barlow-unt…
69172  http://www.tate.org.uk/art/artworks/barlow-unt…
69173  http://www.tate.org.uk/art/artworks/barlow-unt…

[29019 rows x 20 columns]
```

```
[ ]: data.loc[data.title.str.contains('The Circle of the Lustful')]
```

```
[ ]:         id accession_number        artist artistRole  artistId  \
     4      1039           A00005  Blake, William     artist        39
     57394  1156           T01950  Blake, William     artist        39

                                                  title  \
     4      The Circle of the Lustful: Francesca da Rimini…
     57394  The Circle of the Lustful: Francesca da Rimini…

                          dateText                medium  \
     4      1826-7, reprinted 1892  Line engraving on paper
```

```
57394  1826-7, reprinted 1968  Line-engraving on paper

                                              creditLine    year  \
4         Purchased with the assistance of a special gra…  1826.0
57394              Presented by Lessing J. Rosenwald 1975  1826.0

       acquisitionYear          dimensions width height  depth units  \
4              1919.0  image: 243 x 335 mm    243  335.0    NaN    mm
57394          1975.0  image: 243 x 335 mm    243    335    NaN    mm

       inscription thumbnailCopyright  \
4              NaN               NaN
57394          NaN               NaN

                                      thumbnailUrl  \
4      http://www.tate.org.uk/art/images/work/A/A00/A…
57394  http://www.tate.org.uk/art/images/work/T/T01/T…

                                      url
4      http://www.tate.org.uk/art/artworks/blake-the-…
57394  http://www.tate.org.uk/art/artworks/blake-the-…
```

[ ]: