08_indexing_and_filtering

March 21, 2022

1 Indexing and filtering data

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

```python
import pandas as pd
data= pd.read_csv('./data/artwork_sample.csv')
```

```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
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
```

```
        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

    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['id']
```

```
0    1035
1    1036
2    1037
3    1038
4    1039
5    1040
6    1041
7    1042
8    1043
9    1044
Name: id, dtype: int64
```

```python
data['id'][1]
```

```
1036
```

```python
data[['id', 'title']]
```

```
     id                                              title
0  1035  A Figure Bowing before a Seated Old Man with h…
1  1036  Two Drawings of Frightened Figures, Probably f…
2  1037  The Preaching of Warning. Verso: An Old Man En…
3  1038     Six Drawings of Figures with Outstretched Arms
4  1039  The Circle of the Lustful: Francesca da Rimini…
```

```
5   1040          Ciampolo the Barrator Tormented by the Devils
6   1041                           The Baffled Devils Fighting
7   1042   The Six-Footed Serpent Attacking Agnolo Brunel…
8   1043                     The Serpent Attacking Buoso Donati
9   1044                     The Pit of Disease: The Falsifiers
```

[ ]: `data[['artist', 'title']]`

```
[ ]:            artist                                             title
     0   Blake, Robert   A Figure Bowing before a Seated Old Man with h…
     1   Blake, Robert   Two Drawings of Frightened Figures, Probably f…
     2   Blake, Robert   The Preaching of Warning. Verso: An Old Man En…
     3   Blake, Robert      Six Drawings of Figures with Outstretched Arms
     4  Blake, William   The Circle of the Lustful: Francesca da Rimini…
     5  Blake, William       Ciampolo the Barrator Tormented by the Devils
     6  Blake, William                           The Baffled Devils Fighting
     7  Blake, William   The Six-Footed Serpent Attacking Agnolo Brunel…
     8  Blake, William                     The Serpent Attacking Buoso Donati
     9  Blake, William                     The Pit of Disease: The Falsifiers
```

[ ]: `data[['artist', 'title']]['artist']`

```
[ ]: 0       Blake, Robert
     1       Blake, Robert
     2       Blake, Robert
     3       Blake, Robert
     4      Blake, William
     5      Blake, William
     6      Blake, William
     7      Blake, William
     8      Blake, William
     9      Blake, William
     Name: artist, dtype: object
```

[ ]: `data[['artist', 'title']]['artist'][0]`

[ ]: `'Blake, Robert'`

[ ]:
```python
# Accessing direct row index not allowed in DataFrame
data[1]
```

```
---------------------------------------------------------------------------
KeyError                                  Traceback (most recent call last)
File C:\Python310\lib\site-packages\pandas\core\indexes\base.py:3621, in Index.
  ↪get_loc(self, key, method, tolerance)
   3620 try:
-> 3621     return self._engine.get_loc(casted_key)
   3622 except KeyError as err:
```

```
File C:\Python310\lib\site-packages\pandas\_libs\index.pyx:136, in pandas._libs
 ↪index.IndexEngine.get_loc()

File C:\Python310\lib\site-packages\pandas\_libs\index.pyx:163, in pandas._libs
 ↪index.IndexEngine.get_loc()

File pandas\_libs\hashtable_class_helper.pxi:5198, in pandas._libs.hashtable.
 ↪PyObjectHashTable.get_item()

File pandas\_libs\hashtable_class_helper.pxi:5206, in pandas._libs.hashtable.
 ↪PyObjectHashTable.get_item()

KeyError: 1

The above exception was the direct cause of the following exception:

KeyError                                  Traceback (most recent call last)
Input In [10], in <cell line: 2>()
      1 # Accessing direct row index not allowed in DataFrame
----> 2 data[1]

File C:\Python310\lib\site-packages\pandas\core\frame.py:3505, in DataFrame.
 ↪__getitem__(self, key)
   3503 if self.columns.nlevels > 1:
   3504     return self._getitem_multilevel(key)
-> 3505 indexer = self.columns.get_loc(key)
   3506 if is_integer(indexer):
   3507     indexer = [indexer]

File C:\Python310\lib\site-packages\pandas\core\indexes\base.py:3623, in Index.
 ↪get_loc(self, key, method, tolerance)
   3621     return self._engine.get_loc(casted_key)
   3622 except KeyError as err:
-> 3623     raise KeyError(key) from err
   3624 except TypeError:
   3625     # If we have a listlike key, _check_indexing_error will raise
   3626     #  InvalidIndexError. Otherwise we fall through and re-raise
   3627     #  the TypeError.
   3628     self._check_indexing_error(key)

KeyError: 1
```

```
[ ]: # But we can access by range
     data[1:2]
```

```
[ ]:      id accession_number         artist artistRole  artistId  \
     1  1036           A00002  Blake, Robert     artist        38

                                                  title       dateText  \
     1  Two Drawings of Frightened Figures, Probably f…  date not known

                  medium                          creditLine  year  \
     1  Graphite on paper  Presented by Mrs John Richmond 1922   NaN

        acquisitionYear            dimensions  width  height  depth units  \
     1             1922  support: 311 x 213 mm  311.0   213.0    NaN    mm

        inscription  thumbnailCopyright  \
     1          NaN                 NaN

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

                                               url
     1  http://www.tate.org.uk/art/artworks/blake-two-…

[ ]:  # range uses integer position of the rows, but not the row label
      # in this dataset, both are same, but it might not be the case every time.
      data[1:5]

[ ]:      id accession_number         artist artistRole  artistId  \
     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  \
     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  \
     1                     Graphite on paper
     2  Graphite on paper. Verso: graphite on paper
     3                     Graphite on paper
     4                 Line engraving on paper

                                  creditLine    year  acquisitionYear  \
     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
```

```
            dimensions  width  height  depth units  inscription  \
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
```

```
  thumbnailCopyright                              thumbnailUrl  \
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
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[1:5]['artist']
```

```
1      Blake, Robert
2      Blake, Robert
3      Blake, Robert
4    Blake, William
Name: artist, dtype: object
```

```python
# every row of data where year is > 1800 -> then year series
data[data['year'] > 1800]['year']
```

```
4    1826.0
5    1826.0
6    1826.0
7    1826.0
8    1826.0
9    1826.0
Name: year, dtype: float64
```

2 Using .loc() method

```python
# data.loc[ROWS, COLS]
```

```python
# 0th row
data.loc[0]
```

```
[ ]: id                                                 1035
     accession_number                              A00001
     artist                                  Blake, Robert
     artistRole                                     artist
     artistId                                           38
     title            A Figure Bowing before a Seated Old Man with h…
     dateText                                 date not known
     medium           Watercolour, ink, chalk and graphite on paper…
     creditLine              Presented by Mrs John Richmond 1922
     year                                              NaN
     acquisitionYear                                   1922
     dimensions                         support: 394 x 419 mm
     width                                            394.0
     height                                           419.0
     depth                                              NaN
     units                                               mm
     inscription                                        NaN
     thumbnailCopyright                                 NaN
     thumbnailUrl     http://www.tate.org.uk/art/images/work/A/A00/A…
     url              http://www.tate.org.uk/art/artworks/blake-a-fi…
     Name: 0, dtype: object
```

```
[ ]: # 0th row and all columns
     data.loc[0, :]
```

```
[ ]: id                                                 1035
     accession_number                              A00001
     artist                                  Blake, Robert
     artistRole                                     artist
     artistId                                           38
     title            A Figure Bowing before a Seated Old Man with h…
     dateText                                 date not known
     medium           Watercolour, ink, chalk and graphite on paper…
     creditLine              Presented by Mrs John Richmond 1922
     year                                              NaN
     acquisitionYear                                   1922
     dimensions                         support: 394 x 419 mm
     width                                            394.0
     height                                           419.0
     depth                                              NaN
     units                                               mm
     inscription                                        NaN
     thumbnailCopyright                                 NaN
     thumbnailUrl     http://www.tate.org.uk/art/images/work/A/A00/A…
     url              http://www.tate.org.uk/art/artworks/blake-a-fi…
     Name: 0, dtype: object
```

```
# 0th row and specific columns
data.loc[0:, ['title']]
```

```
                                              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…
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
9                     The Pit of Disease: The Falsifiers
```

```
# specific conditions and specific columns
data.loc[data['year']>1800, ['title', 'year']]
```

```
                                              title     year
4  The Circle of the Lustful: Francesca da Rimini…  1826.0
5       Ciampolo the Barrator Tormented by the Devils  1826.0
6                           The Baffled Devils Fighting  1826.0
7  The Six-Footed Serpent Attacking Agnolo Brunel…  1826.0
8                     The Serpent Attacking Buoso Donati  1826.0
9                     The Pit of Disease: The Falsifiers  1826.0
```

```
# specific row labels (only where row label is 0 and 5). Its label NOT POSITION
data.loc[[1,5], ['title', 'year']]
```

```
                                              title     year
1  Two Drawings of Frightened Figures, Probably f…    NaN
5       Ciampolo the Barrator Tormented by the Devils  1826.0
```

```
# specific rows and range of columns
data.loc[[1,5], 'id': 'artistId']
```

```
     id accession_number         artist artistRole  artistId
1  1036           A00002   Blake, Robert     artist        38
5  1040           A00006  Blake, William     artist        39
```

```
data.loc[data['artist'] == 'Blake, Robert', 'id':'artistId']
```

```
     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
```

3 Using iloc()

```python
# iloc uses integer positions of the rows and columns instead of labels
data.iloc[0, :]
```

```
id                                               1035
accession_number                               A00001
artist                                  Blake, Robert
artistRole                                     artist
artistId                                           38
title               A Figure Bowing before a Seated Old Man with h…
dateText                                date not known
medium              Watercolour, ink, chalk and graphite on paper…
creditLine                  Presented by Mrs John Richmond 1922
year                                              NaN
acquisitionYear                                  1922
dimensions                         support: 394 x 419 mm
width                                           394.0
height                                          419.0
depth                                             NaN
units                                              mm
inscription                                       NaN
thumbnailCopyright                                NaN
thumbnailUrl        http://www.tate.org.uk/art/images/work/A/A00/A…
url                 http://www.tate.org.uk/art/artworks/blake-a-fi…
Name: 0, dtype: object
```

```python
# since it uses integer positions on rows and columns, we can specify
# column integer range as well
data.iloc[0:3, 0:5]
```

```
     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
```

```python
df = data.copy()
# To avoid confusion, lets set the index as id
df.set_index('id', inplace=True)
```

```python
# returns rows because these integers are position numbers
df.iloc[0:3, 0:5]
```

```
     accession_number         artist artistRole  artistId  \
id
1035           A00001  Blake, Robert     artist        38
1036           A00002  Blake, Robert     artist        38
```

```
1037             A00003  Blake, Robert       artist         38

                                                  title
id
1035  A Figure Bowing before a Seated Old Man with h…
1036  Two Drawings of Frightened Figures, Probably f…
1037  The Preaching of Warning. Verso: An Old Man En…
```

```
# returns nothing, because now the df does not contain 0-3 labels in index
# instead now index contains 1035 type of numbers
df.loc[0:3, :]
```

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

```
df.loc[1035:1037, :]
```

```
     accession_number        artist artistRole  artistId  \
id
1035           A00001  Blake, Robert     artist        38
1036           A00002  Blake, Robert     artist        38
1037           A00003  Blake, Robert     artist        38

                                                  title        dateText  \
id
1035  A Figure Bowing before a Seated Old Man with h…  date not known
1036  Two Drawings of Frightened Figures, Probably f…  date not known
1037  The Preaching of Warning. Verso: An Old Man En…         ?c.1785

                                         medium  \
id
1035  Watercolour, ink, chalk and graphite on paper…
1036                              Graphite on paper
1037      Graphite on paper. Verso: graphite on paper

                             creditLine    year  acquisitionYear  \
id
1035  Presented by Mrs John Richmond 1922    NaN             1922
1036  Presented by Mrs John Richmond 1922    NaN             1922
1037  Presented by Mrs John Richmond 1922  1785.0             1922

                 dimensions  width  height  depth units  inscription  \
id
1035  support: 394 x 419 mm  394.0   419.0    NaN    mm          NaN
```

```
1036  support: 311 x 213 mm  311.0   213.0   NaN    mm          NaN
1037  support: 343 x 467 mm  343.0   467.0   NaN    mm          NaN

        thumbnailCopyright                               thumbnailUrl  \
id
1035                 NaN   http://www.tate.org.uk/art/images/work/A/A00/A…
1036                 NaN   http://www.tate.org.uk/art/images/work/A/A00/A…
1037                 NaN   http://www.tate.org.uk/art/images/work/A/A00/A…

                                                      url
id
1035   http://www.tate.org.uk/art/artworks/blake-a-fi…
1036   http://www.tate.org.uk/art/artworks/blake-two-…
1037   http://www.tate.org.uk/art/artworks/blake-the-…
```

```python
# now id is index, so column 0 is accession_number and 2 is artistRole, 3rd is
↪exclusive in indexing
df.iloc[0:3, 0:3]
```

```
      accession_number         artist artistRole
id
1035           A00001  Blake, Robert     artist
1036           A00002  Blake, Robert     artist
1037           A00003  Blake, Robert     artist
```

```python
#lets reset the index to check it
df.reset_index(inplace=True)
```

```python
# now as the index has been reset, so 0 column becomes the id, and 3 becomes
↪the artistRole which is exclusive
# and we are getting 0,1, and 2 index i.e. id, accession_number, and artist
df.iloc[0:3, 0:3]
```

```
     id accession_number         artist
0  1035           A00001  Blake, Robert
1  1036           A00002  Blake, Robert
2  1037           A00003  Blake, Robert
```

```python
# QUESTION: 5, 8 should return row number 4,7, why it is returning 5,8
# while columns are returning correct index
df.iloc[[5,8], [0, 2]]
```

```
     id          artist
5  1040  Blake, William
8  1043  Blake, William
```

```python
df.iloc[0:8, [0, 2]]
```
```

```
[ ]:      id        artist
     0  1035    Blake, Robert
     1  1036    Blake, Robert
     2  1037    Blake, Robert
     3  1038    Blake, Robert
     4  1039  Blake, William
     5  1040  Blake, William
     6  1041  Blake, William
     7  1042  Blake, William
```

```
[ ]: df.iloc[5:8, [0, 2]]
```

4 Filtering Data

4.0.1 str.contains

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

```
             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


    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.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
Name: medium, dtype: object
```

[ ]: ```python
type(data.medium)
```

[ ]: ```
pandas.core.series.Series
```

[ ]: ```python
data.medium.str
```

[ ]: ```
<pandas.core.strings.accessor.StringMethods at 0x2b235060880>
```

[ ]: ```python
data.medium.str.contains('Graphite')
```

[ ]: ```
0   False
1   True
2   True
```

```
3     True
4     False
5     False
6     False
7     False
8     False
9     False
Name: medium, dtype: bool
```

[ ]: ```
# data.medium.str.lower().contains('graphite')
data.medium.str.lower().str.contains('graphite')
```

[ ]: ```
0   True
1   True
2   True
3   True
4   False
5   False
6   False
7   False
8   False
9   False
Name: medium, dtype: bool
```

[ ]: ```
data.loc[data.medium.str.contains('Graphite'), ['artist', 'medium']]
```

[ ]:
	artist	medium
1	Blake, Robert	Graphite on paper
2	Blake, Robert	Graphite on paper. Verso: graphite on paper
3	Blake, Robert	Graphite on paper

[ ]: ```
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-…
```

```python
data.loc[data.medium.str.contains('Graphite', case=False), ['artist', 'medium']]
```

```
            artist                                             medium
0  Blake, Robert  Watercolour, ink, chalk and graphite on paper…
1  Blake, Robert                                  Graphite on paper
2  Blake, Robert         Graphite on paper. Verso: graphite on paper
3  Blake, Robert                                  Graphite on paper
```

```python
data.loc[data.medium.str.contains('(?i)Graphite', regex=True) | data.medium.str.
 contains('(?i)Line', regex=True), ['artist', 'medium']]
```

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

```
data.loc[data.medium.str.contains('graphite|line', case=False, regex=True),
         ['artist', 'medium', 'title']]
```

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

                                              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…
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
9               The Pit of Disease: The Falsifiers
```

```
data.dtypes
```

```
id                      int64
accession_number       object
artist                 object
artistRole             object
artistId                int64
title                  object
dateText               object
medium                 object
creditLine             object
year                  float64
acquisitionYear         int64
dimensions             object
width                 float64
height                float64
depth                 float64
units                  object
inscription           float64
thumbnailCopyright    float64
```

```
thumbnailUrl           object
url                    object
dtype: object
```

`[ ]:` `data.year.str.contains('1826')`

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
Input In [69], in <cell line: 1>()
----> 1 data.year.str.contains('1826')

File C:\Python310\lib\site-packages\pandas\core\generic.py:5583, in NDFrame.
 ↪__getattr__(self, name)
   5576 if (
   5577     name not in self._internal_names_set
   5578     and name not in self._metadata
   5579     and name not in self._accessors
   5580     and self._info_axis._can_hold_identifiers_and_holds_name(name)
   5581 ):
   5582     return self[name]
-> 5583 return object.__getattribute__(self, name)

File C:\Python310\lib\site-packages\pandas\core\accessor.py:182, in
 ↪CachedAccessor.__get__(self, obj, cls)
    179 if obj is None:
    180     # we're accessing the attribute of the class, i.e., Dataset.geo
    181     return self._accessor
--> 182 accessor_obj = self._accessor(obj)
    183 # Replace the property with the accessor object. Inspired by:
    184 # https://www.pydanny.com/cached-property.html
    185 # We need to use object.__setattr__ because we overwrite __setattr__ on
    186 # NDFrame
    187 object.__setattr__(obj, self._name, accessor_obj)

File C:\Python310\lib\site-packages\pandas\core\strings\accessor.py:177, in
 ↪StringMethods.__init__(self, data)
    174 def __init__(self, data):
    175     from pandas.core.arrays.string_ import StringDtype
--> 177     self._inferred_dtype = self._validate(data)
    178     self._is_categorical = is_categorical_dtype(data.dtype)
    179     self._is_string = isinstance(data.dtype, StringDtype)

File C:\Python310\lib\site-packages\pandas\core\strings\accessor.py:231, in
 ↪StringMethods._validate(data)
    228 inferred_dtype = lib.infer_dtype(values, skipna=True)
    230 if inferred_dtype not in allowed_types:
```

```
--> 231        raise AttributeError("Can only use .str accessor with string values
    ↪")
     232 return inferred_dtype

AttributeError: Can only use .str accessor with string values!
```

```
[ ]: data.year.astype(str).str.contains('1826')
```

```
[ ]: 0     False
     1     False
     2     False
     3     False
     4      True
     5      True
     6      True
     7      True
     8      True
     9      True
     Name: year, dtype: bool
```

```
[ ]: data.loc[data.year.astype(str).str.contains('1826'), ['title', 'year']]
```

```
[ ]:                                                  title    year
     4  The Circle of the Lustful: Francesca da Rimini…  1826.0
     5      Ciampolo the Barrator Tormented by the Devils  1826.0
     6                         The Baffled Devils Fighting  1826.0
     7  The Six-Footed Serpent Attacking Agnolo Brunel…  1826.0
     8                 The Serpent Attacking Buoso Donati  1826.0
     9                 The Pit of Disease: The Falsifiers  1826.0
```

```
[ ]: 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
data.loc[data.dimensions.str.contains('support')]
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
Input In [75], in <cell line: 1>()
----> 1 data.loc[data.dimensions.str.contains('support')]

File C:\Python310\lib\site-packages\pandas\core\indexing.py:967, in
  ↪_LocationIndexer.__getitem__(self, key)
    964 axis = self.axis or 0
    966 maybe_callable = com.apply_if_callable(key, self.obj)
```

```
--> 967 return self._getitem_axis(maybe_callable, axis=axis)

File C:\Python310\lib\site-packages\pandas\core\indexing.py:1181, in _LocIndexer.
  ↪_getitem_axis(self, key, axis)
   1179        self._validate_key(key, axis)
   1180        return self._get_slice_axis(key, axis=axis)
-> 1181 elif com.is_bool_indexer(key):
   1182        return self._getbool_axis(key, axis=axis)
   1183 elif is_list_like_indexer(key):
   1184
   1185        # an iterable multi-selection

File C:\Python310\lib\site-packages\pandas\core\common.py:144, in
  ↪is_bool_indexer(key)
    140        na_msg = "Cannot mask with non-boolean array containing NA / NaN
  ↪values"
    141        if lib.infer_dtype(key) == "boolean" and isna(key).any():
    142            # Don't raise on e.g. ["A", "B", np.nan], see
    143            #  test_loc_getitem_list_of_labels_categoricalindex_with_na
--> 144            raise ValueError(na_msg)
    145        return False
    146 return True

ValueError: Cannot mask with non-boolean array containing NA / NaN values
```

```
[ ]: data.loc[data.dimensions.str.contains('support', na=False),␣
     ↪['title','dimensions', 'year',]]
```

```
[ ]:                                                      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
     32         Sketch for 'The Four and Twenty Elders Casting…
     …                                                        …
     69170                                             Untitled
     69171                                             Untitled
     69172                                             Untitled
     69173                                             Untitled
     69178                           Will We Ever Kill The Bug

                         dimensions  year
     0        support: 394 x 419 mm   NaN
     1        support: 311 x 213 mm   NaN
     2        support: 343 x 467 mm   1785
     3        support: 318 x 394 mm   NaN
     32       support: 488 x 389 mm   1803
```

```
 …                               …     …
69170             support: 560 x 762 mm   2003
69171             support: 558 x 760 mm   2003
69172             support: 380 x 566 mm   2004
69173             support: 560 x 762 mm   2006
69178   support, each: 438 x 330 mm   2010

[49736 rows x 3 columns]
```

[ ]: