02_aggregate

March 21, 2022

1 Aggregate 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-…
```

[ ]: `data.year`

[ ]:
```
0       NaN
1       NaN
2    1785.0
3       NaN
4    1826.0
5    1826.0
6    1826.0
7    1826.0
8    1826.0
9    1826.0
Name: year, dtype: float64
```

[ ]:
```
# nan values are automatically excluded from aggregate functions
data.year.min()
```

[ ]: `1785.0`

[ ]: `data.year.max()`

[ ]: `1826.0`

[ ]: `data.year.sum()`

[ ]: `12741.0`

```
[ ]: data.year.mean()
```

```
[ ]: 1820.142857142857
```

```
[ ]: # standard deviation
     data.year.std()
```

```
[ ]: 15.496543393378316
```

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

```
[ ]: 'Blake, Robert'
```

```
[ ]: # mean of string values does not makes sense, so it will return error
     data.artist.mean()
```

```
    ---------------------------------------------------------------------------
    ValueError                                Traceback (most recent call last)
    File C:\Python310\lib\site-packages\pandas\core\nanops.py:1622, in
     ↪_ensure_numeric(x)
       1621 try:
    -> 1622     x = float(x)
       1623 except (TypeError, ValueError):
       1624     # e.g. "1+1j" or "foo"

    ValueError: could not convert string to float: 'Blake, RobertBlake, RobertBlake
     ↪RobertBlake, RobertBlake, WilliamBlake, WilliamBlake, WilliamBlake,
     ↪WilliamBlake, WilliamBlake, William'

    During handling of the above exception, another exception occurred:

    ValueError                                Traceback (most recent call last)
    File C:\Python310\lib\site-packages\pandas\core\nanops.py:1626, in
     ↪_ensure_numeric(x)
```

```
   1625 try:
-> 1626     x = complex(x)
   1627 except ValueError as err:
   1628     # e.g. "foo"

ValueError: complex() arg is a malformed string

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

TypeError                                 Traceback (most recent call last)
Input In [13], in <cell line: 2>()
      1 # mean of string values does not makes sense, so it will return error
----> 2 data.artist.mean()

File C:\Python310\lib\site-packages\pandas\core\generic.py:11127, in NDFrame.
 _add_numeric_operations.<locals>.mean(self, axis, skipna, level, numeric_only
 **kwargs)
  11109 @doc(
  11110     _num_doc,
  11111     desc="Return the mean of the values over the requested axis.",
  (…)
  11125     **kwargs,
  11126 ):
> 11127     return NDFrame.mean(self, axis, skipna, level, numeric_only,
 **kwargs)

File C:\Python310\lib\site-packages\pandas\core\generic.py:10697, in NDFrame.
 mean(self, axis, skipna, level, numeric_only, **kwargs)
  10689 def mean(
  10690     self,
  10691     axis: Axis | None | lib.NoDefault = lib.no_default,
  (…)
  10695     **kwargs,
  10696 ) -> Series | float:
> 10697     return self._stat_function(
  10698         "mean", nanops.nanmean, axis, skipna, level, numeric_only,
 **kwargs
  10699     )

File C:\Python310\lib\site-packages\pandas\core\generic.py:10649, in NDFrame.
 _stat_function(self, name, func, axis, skipna, level, numeric_only, **kwargs)
  10639     warnings.warn(
  10640         "Using the level keyword in DataFrame and Series aggregations i
 "
  10641         "deprecated and will be removed in a future version. Use groupb
 "
  (…)
  10644         stacklevel=find_stack_level(),
```

```
  10645        )
  10646        return self._agg_by_level(
  10647            name, axis=axis, level=level, skipna=skipna,␣
  ↪numeric_only=numeric_only
  10648        )
> 10649   return self._reduce(
  10650        func, name=name, axis=axis, skipna=skipna, numeric_only=numeric_only
  10651   )

File C:\Python310\lib\site-packages\pandas\core\series.py:4471, in Series.
  ↪_reduce(self, op, name, axis, skipna, numeric_only, filter_type, **kwds)
   4467        raise NotImplementedError(
   4468            f"Series.{name} does not implement {kwd_name}."
   4469        )
   4470    with np.errstate(all="ignore"):
-> 4471        return op(delegate, skipna=skipna, **kwds)

File C:\Python310\lib\site-packages\pandas\core\nanops.py:93, in disallow.
  ↪__call__.<locals>._f(*args, **kwargs)
     91 try:
     92     with np.errstate(invalid="ignore"):
---> 93         return f(*args, **kwargs)
     94 except ValueError as e:
     95     # we want to transform an object array
     96     # ValueError message to the more typical TypeError
     97     # e.g. this is normally a disallowed function on
     98     # object arrays that contain strings
     99     if is_object_dtype(args[0]):

File C:\Python310\lib\site-packages\pandas\core\nanops.py:155, in␣
  ↪bottleneck_switch.__call__.<locals>.f(values, axis, skipna, **kwds)
    153         result = alt(values, axis=axis, skipna=skipna, **kwds)
    154 else:
--> 155     result = alt(values, axis=axis, skipna=skipna, **kwds)
    157 return result

File C:\Python310\lib\site-packages\pandas\core\nanops.py:410, in␣
  ↪_datetimelike_compat.<locals>.new_func(values, axis, skipna, mask, **kwargs)
    407 if datetimelike and mask is None:
    408     mask = isna(values)
--> 410 result = func(values, axis=axis, skipna=skipna, mask=mask, **kwargs)
    412 if datetimelike:
    413     result = _wrap_results(result, orig_values.dtype, fill_value=iNaT)

File C:\Python310\lib\site-packages\pandas\core\nanops.py:698, in␣
  ↪nanmean(values, axis, skipna, mask)
    695     dtype_count = dtype
    697 count = _get_counts(values.shape, mask, axis, dtype=dtype_count)
```

```
--> 698 the_sum = _ensure_numeric(values.sum(axis, dtype=dtype_sum))
    700 if axis is not None and getattr(the_sum, "ndim", False):
    701     count = cast(np.ndarray, count)

File C:\Python310\lib\site-packages\pandas\core\nanops.py:1629, in
  ↪_ensure_numeric(x)
   1626             x = complex(x)
   1627         except ValueError as err:
   1628             # e.g. "foo"
-> 1629             raise TypeError(f"Could not convert {x} to numeric") from err
   1630     return x

TypeError: Could not convert Blake, RobertBlake, RobertBlake, RobertBlake,
  ↪RobertBlake, WilliamBlake, WilliamBlake, WilliamBlake, WilliamBlake,
  ↪WilliamBlake, William to numeric
```

```python
# sum on string concatenates the values
data.artist.sum()
```

```
'Blake, RobertBlake, RobertBlake, RobertBlake, RobertBlake, WilliamBlake,
WilliamBlake, WilliamBlake, WilliamBlake, WilliamBlake, William'
```

```python
# sum can be called on entire data frame instead of a single series too
# Though the result is not what we want.
data.sum()
```

```
id                                                         10395
accession_number    A00001A00002A00003A00004A00005A00006A00007A000…
artist              Blake, RobertBlake, RobertBlake, RobertBlake, …
artistRole          artistartistartistartistartistartistartistarti…
artistId                                                     386
title               A Figure Bowing before a Seated Old Man with h…
dateText            date not knowndate not known?c.1785date not kn…
medium              Watercolour, ink, chalk and graphite on paper…
creditLine          Presented by Mrs John Richmond 1922Presented b…
year                                                     12741.0
acquisitionYear                                            19202
dimensions          support: 394 x 419 mmsupport: 311 x 213 mmsupp…
width                                                     2821.0
height                                                    3515.0
depth                                                        0.0
units               mmmmmmmmmmmmmmmmmmmmmmmmm
inscription                                                  0.0
thumbnailCopyright                                           0.0
thumbnailUrl        http://www.tate.org.uk/art/images/work/A/A00/A…
url                 http://www.tate.org.uk/art/artworks/blake-a-fi…
dtype: object
```

```
data.agg('min')
```

```
id                                                1035
accession_number                                A00001
artist                                    Blake, Robert
artistRole                                       artist
artistId                                             38
title             A Figure Bowing before a Seated Old Man with h…
dateText                           1826-7, reprinted 1892
medium                                  Graphite on paper
creditLine              Presented by Mrs John Richmond 1922
year                                             1785.0
acquisitionYear                                    1919
dimensions                           image: 240 x 338 mm
width                                            240.0
height                                           213.0
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…
dtype: object
```

```
# more than one aggregations at the same time
data.agg(['min', 'max'])
```

```
         id accession_number            artist artistRole  artistId  \
min    1035           A00001     Blake, Robert     artist        38
max    1044           A00010    Blake, William     artist        39

                                              title  \
min  A Figure Bowing before a Seated Old Man with h…
max  Two Drawings of Frightened Figures, Probably f…

                     dateText  \
min  1826-7, reprinted 1892
max          date not known

                                            medium  \
min                          Graphite on paper
max  Watercolour, ink, chalk and graphite on paper…

                                   creditLine     year  \
min             Presented by Mrs John Richmond 1922  1785.0
max  Purchased with the assistance of a special gra…  1826.0
```

```
       acquisitionYear              dimensions  width  height  depth units  \
min             1919     image: 240 x 338 mm  240.0   213.0   NaN    mm
max             1922  support: 394 x 419 mm  394.0   467.0   NaN    mm


     inscription  thumbnailCopyright  \
min          NaN                 NaN
max          NaN                 NaN


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


                                                  url
min  http://www.tate.org.uk/art/artworks/blake-a-fi…
max  http://www.tate.org.uk/art/artworks/blake-two-…
```

[ ]: `data.agg(['min', 'max', 'mean', 'std'])`

```
C:\Users\hp\AppData\Local\Temp\ipykernel_8068\3964390368.py:1: FutureWarning:
['accession_number', 'artist', 'artistRole', 'title', 'dateText', 'medium',
'creditLine', 'dimensions', 'units', 'thumbnailUrl', 'url'] did not aggregate
successfully. If any error is raised this will raise in a future version of
pandas. Drop these columns/ops to avoid this warning.
  data.agg(['min', 'max', 'mean', 'std'])
```

```
[ ]:             id accession_number         artist artistRole    artistId  \
     min   1035.00000           A00001  Blake, Robert     artist   38.000000
     max   1044.00000           A00010  Blake, William     artist   39.000000
     mean  1039.50000              NaN            NaN        NaN   38.600000
     std      3.02765              NaN            NaN        NaN    0.516398


                                                 title  \
     min   A Figure Bowing before a Seated Old Man with h…
     max   Two Drawings of Frightened Figures, Probably f…
     mean                                            NaN
     std                                             NaN


                       dateText  \
     min   1826-7, reprinted 1892
     max          date not known
     mean                     NaN
     std                      NaN


                                                 medium  \
     min                              Graphite on paper
     max   Watercolour, ink, chalk and graphite on paper…
     mean                                           NaN
     std                                            NaN
```

```
                                            creditLine           year  \
min                       Presented by Mrs John Richmond 1922   1785.000000
max     Purchased with the assistance of a special gra…   1826.000000
mean                                                NaN   1820.142857
std                                                 NaN     15.496543

        acquisitionYear            dimensions      width      height  depth  \
min         1919.000000    image: 240 x 338 mm  240.000000  213.000000    NaN
max         1922.000000  support: 394 x 419 mm  394.000000  467.000000    NaN
mean        1920.200000                    NaN  282.100000  351.500000    NaN
std            1.549193                    NaN   55.562677   66.818577    NaN

        units  inscription  thumbnailCopyright  \
min        mm          NaN                 NaN
max        mm          NaN                 NaN
mean      NaN          NaN                 NaN
std       NaN          NaN                 NaN

                              thumbnailUrl  \
min    http://www.tate.org.uk/art/images/work/A/A00/A…
max    http://www.tate.org.uk/art/images/work/A/A00/A…
mean                                           NaN
std                                            NaN

                                      url
min    http://www.tate.org.uk/art/artworks/blake-a-fi…
max    http://www.tate.org.uk/art/artworks/blake-two-…
mean                                           NaN
std                                            NaN
```

```python
# aggregation can be used on columns as well
data.agg('mean', axis=1)
```

```
C:\Users\hp\AppData\Local\Temp\ipykernel_8068\849437883.py:2: FutureWarning:
Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None')
is deprecated; in a future version this will raise TypeError.  Select only valid
columns before calling the reduction.
  data.agg('mean', axis=1)
```

```
Series([], dtype: float64)
```