In [36]: ▶|
```python
#Paul Galvez
#Class: DSC 540
#Date: 5/4/23
```

In [37]: ▶|
```python
#Loading the libraries needed for the project

import pandas as pd
import numpy as np
```

In [38]: ▶|
```python
#Loading the CSV data

df = pd.read_csv("MetObjects.csv")
df
```

C:\Users\paul_\AppData\Local\Temp\ipykernel_21472\1752111644.py:1: DtypeWarning: Columns (5,7,10,11,12,1
3,14,34,35,36,37,38,39,40,41,42,43,44,45,46) have mixed types. Specify dtype option on import or set low
_memory=False.
  df = pd.read_csv("MetObjects.csv")

Out[38]:

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
0	1979.486.1	False	False	False	1	NaN	The American Wing	1979.0	Coin	One-dollar Liberty Head Coin	...	NaN	
1	1980.264.5	False	False	False	2	NaN	The American Wing	1980.0	Coin	Ten-dollar Liberty Head Coin	...	NaN	
2	67.265.9	False	False	False	3	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
3	67.265.10	False	False	False	4	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
4	67.265.11	False	False	False	5	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
477799	23.112.2893	False	False	True	860869	NaN	Drawings and Prints	1923	Drawing	Phaeton #24567	...	NaN	
477800	23.112.2894	False	False	True	860870	NaN	Drawings and Prints	1923	Drawing	Stanhope Phaeton #25538-25539 (#21222)	...	NaN	
477801	53.600.1434	False	False	False	860871	NaN	Drawings and Prints	1953	Print	Forest landscape with cattle drinking, a woman...	...	NaN	

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
477802	23.112.2895	False	False	True	860872	NaN	Drawings and Prints	1923	Drawing	Phaeton with folding top	...	NaN	
477803	23.112.2896	False	False	True	860873	NaN	Drawings and Prints	1923	Drawing	Stanhope Phaeton	...	NaN	

477804 rows × 54 columns

In [39]: ▶| `df.head()`

Out[39]:

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Classification
0	1979.486.1	False	False	False	1	NaN	The American Wing	1979.0	Coin	One-dollar Liberty Head Coin	...	NaN	NaN
1	1980.264.5	False	False	False	2	NaN	The American Wing	1980.0	Coin	Ten-dollar Liberty Head Coin	...	NaN	NaN
2	67.265.9	False	False	False	3	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	NaN
3	67.265.10	False	False	False	4	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	NaN
4	67.265.11	False	False	False	5	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	NaN

5 rows × 54 columns

◀ ▬▬▬▬▬▬▬▬▬▬▬▬▬▬ ▶

```
In [40]:  ▶ df.columns
```

```
Out[40]: Index(['Object Number', 'Is Highlight', 'Is Timeline Work', 'Is Public Domain',
                'Object ID', 'Gallery Number', 'Department', 'AccessionYear',
                'Object Name', 'Title', 'Culture', 'Period', 'Dynasty', 'Reign',
                'Portfolio', 'Constituent ID', 'Artist Role', 'Artist Prefix',
                'Artist Display Name', 'Artist Display Bio', 'Artist Suffix',
                'Artist Alpha Sort', 'Artist Nationality', 'Artist Begin Date',
                'Artist End Date', 'Artist Gender', 'Artist ULAN URL',
                'Artist Wikidata URL', 'Object Date', 'Object Begin Date',
                'Object End Date', 'Medium', 'Dimensions', 'Credit Line',
                'Geography Type', 'City', 'State', 'County', 'Country', 'Region',
                'Subregion', 'Locale', 'Locus', 'Excavation', 'River', 'Classification',
                'Rights and Reproduction', 'Link Resource', 'Object Wikidata URL',
                'Metadata Date', 'Repository', 'Tags', 'Tags AAT URL',
                'Tags Wikidata URL'],
               dtype='object')
```

```
In [41]:   df.describe
           df
```

Out[41]:

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
0	1979.486.1	False	False	False	1	NaN	The American Wing	1979.0	Coin	One-dollar Liberty Head Coin	...	NaN	
1	1980.264.5	False	False	False	2	NaN	The American Wing	1980.0	Coin	Ten-dollar Liberty Head Coin	...	NaN	
2	67.265.9	False	False	False	3	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
3	67.265.10	False	False	False	4	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
4	67.265.11	False	False	False	5	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	
477799	23.112.2893	False	False	True	860869	NaN	Drawings and Prints	1923	Drawing	Phaeton #24567	...	NaN	
477800	23.112.2894	False	False	True	860870	NaN	Drawings and Prints	1923	Drawing	Stanhope Phaeton #25538-25539 (#21222)	...	NaN	
477801	53.600.1434	False	False	False	860871	NaN	Drawings and Prints	1953	Print	Forest landscape with cattle drinking, a woman...	...	NaN	

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
477802	23.112.2895	False	False	True	860872	NaN	Drawings and Prints	1923	Drawing	Phaeton with folding top	...	NaN	
477803	23.112.2896	False	False	True	860873	NaN	Drawings and Prints	1923	Drawing	Stanhope Phaeton	...	NaN	

477804 rows × 54 columns

In [22]: ▶| `#Chapter 7`

In [25]: ▶| `#Filter out missing data`

`df.dropna()`

Out[25]:

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Classification	Rig Repr

0 rows × 54 columns

In [27]: ▶|
```python
#Filter out missing data

df.dropna(how='all')
```

Out[27]:

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
0	1979.486.1	False	False	False	1	NaN	The American Wing	1979.0	Coin	One-dollar Liberty Head Coin	...	NaN	
1	1980.264.5	False	False	False	2	NaN	The American Wing	1980.0	Coin	Ten-dollar Liberty Head Coin	...	NaN	
2	67.265.9	False	False	False	3	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
3	67.265.10	False	False	False	4	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
4	67.265.11	False	False	False	5	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	
477799	23.112.2893	False	False	True	860869	NaN	Drawings and Prints	1923	Drawing	Phaeton #24567	...	NaN	
477800	23.112.2894	False	False	True	860870	NaN	Drawings and Prints	1923	Drawing	Stanhope Phaeton #25538-25539 (#21222)	...	NaN	
477801	53.600.1434	False	False	False	860871	NaN	Drawings and Prints	1953	Print	Forest landscape with cattle drinking, a woman...	...	NaN	

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
477802	23.112.2895	False	False	True	860872	NaN	Drawings and Prints	1923	Drawing	Phaeton with folding top	...	NaN	
477803	23.112.2896	False	False	True	860873	NaN	Drawings and Prints	1923	Drawing	Stanhope Phaeton	...	NaN	

477804 rows × 54 columns

In [30]:
```python
#Fill in missing data

df.fillna("Filler")
```

Out[30]:

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
0	1979.486.1	False	False	False	1	Filler	The American Wing	1979.0	Coin	One-dollar Liberty Head Coin	...	Filler	
1	1980.264.5	False	False	False	2	Filler	The American Wing	1980.0	Coin	Ten-dollar Liberty Head Coin	...	Filler	
2	67.265.9	False	False	False	3	Filler	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	Filler	
3	67.265.10	False	False	False	4	Filler	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	Filler	
4	67.265.11	False	False	False	5	Filler	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	Filler	
...	...	...	...	...	...	...	...	...	...	...	...	...	
477799	23.112.2893	False	False	True	860869	Filler	Drawings and Prints	1923	Drawing	Phaeton #24567	...	Filler	
477800	23.112.2894	False	False	True	860870	Filler	Drawings and Prints	1923	Drawing	Stanhope Phaeton #25538-25539 (#21222)	...	Filler	
477801	53.600.1434	False	False	False	860871	Filler	Drawings and Prints	1953	Print	Forest landscape with cattle drinking, a woman...	...	Filler	

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
477802	23.112.2895	False	False	True	860872	Filler	Drawings and Prints	1923	Drawing	Phaeton with folding top	...	Filler	
477803	23.112.2896	False	False	True	860873	Filler	Drawings and Prints	1923	Drawing	Stanhope Phaeton	...	Filler	

477804 rows × 54 columns

In [31]: ▶|
```
#Remove duplicates

df.duplicated()
```

Out[31]: 0   False
1   False
2   False
3   False
4   False
          ...
477799   False
477800   False
477801   False
477802   False
477803   False
Length: 477804, dtype: bool

In [32]:

```python
#Remove duplicates

df.drop_duplicates()
```

Out[32]:

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
0	1979.486.1	False	False	False	1	NaN	The American Wing	1979.0	Coin	One-dollar Liberty Head Coin	...	NaN	
1	1980.264.5	False	False	False	2	NaN	The American Wing	1980.0	Coin	Ten-dollar Liberty Head Coin	...	NaN	
2	67.265.9	False	False	False	3	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
3	67.265.10	False	False	False	4	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
4	67.265.11	False	False	False	5	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
477799	23.112.2893	False	False	True	860869	NaN	Drawings and Prints	1923	Drawing	Phaeton #24567	...	NaN	
477800	23.112.2894	False	False	True	860870	NaN	Drawings and Prints	1923	Drawing	Stanhope Phaeton #25538-25539 (#21222)	...	NaN	
477801	53.600.1434	False	False	False	860871	NaN	Drawings and Prints	1953	Print	Forest landscape with cattle drinking, a woman...	...	NaN	

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
477802	23.112.2895	False	False	True	860872	NaN	Drawings and Prints	1923	Drawing	Phaeton with folding top	...	NaN	
477803	23.112.2896	False	False	True	860873	NaN	Drawings and Prints	1923	Drawing	Stanhope Phaeton	...	NaN	

477804 rows × 54 columns

In [35]: ▶

```python
#Chapter 8

df
```

Out[35]:

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
0	1979.486.1	False	False	False	1	NaN	The American Wing	1979.0	Coin	One-dollar Liberty Head Coin	...	NaN	
1	1980.264.5	False	False	False	2	NaN	The American Wing	1980.0	Coin	Ten-dollar Liberty Head Coin	...	NaN	
2	67.265.9	False	False	False	3	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
3	67.265.10	False	False	False	4	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
4	67.265.11	False	False	False	5	NaN	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	...	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	
477799	23.112.2893	False	False	True	860869	NaN	Drawings and Prints	1923	Drawing	Phaeton #24567	...	NaN	
477800	23.112.2894	False	False	True	860870	NaN	Drawings and Prints	1923	Drawing	Stanhope Phaeton #25538-25539 (#21222)	...	NaN	
477801	53.600.1434	False	False	False	860871	NaN	Drawings and Prints	1953	Print	Forest landscape with cattle drinking, a woman...	...	NaN	

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Gallery Number	Department	AccessionYear	Object Name	Title	...	River	Cla
477802	23.112.2895	False	False	True	860872	NaN	Drawings and Prints	1923	Drawing	Phaeton with folding top	...	NaN	
477803	23.112.2896	False	False	True	860873	NaN	Drawings and Prints	1923	Drawing	Stanhope Phaeton	...	NaN	

477804 rows × 54 columns

```
In [34]:  #Create hierarchical index

df.index
```

Out[34]: RangeIndex(start=0, stop=477804, step=1)

In [51]: ▶|
```
#Reshape

my_stack = df.stack()
my_stack
```

Out[51]: 0   Object Number   1979.486.1
             Is Highlight   False
             Is Timeline Work   False
             Is Public Domain   False
             Object ID   1
                                            ...
         477803 Link Resource   http://www.metmuseum.org/art/collection/search... (http://www.metmuseum.or
         g/art/collection/search...)
             Repository   Metropolitan Museum of Art, New York, NY
             Tags   Carriages
             Tags AAT URL   http://vocab.getty.edu/page/aat/300185335 (http://vocab.getty.edu/p
         age/aat/300185335)
             Tags Wikidata URL   https://www.wikidata.org/wiki/Q235356 (https://www.wikidata.or
         g/wiki/Q235356)
         Length: 13015049, dtype: object

In [52]: ▶|
```python
#Reshape

my_stack.unstack()
```

Out[52]:

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Department	AccessionYear	Object Name	Title	Constituent ID	...	Classif
0	1979.486.1	False	False	False	1	The American Wing	1979.0	Coin	One-dollar Liberty Head Coin	16429	...	
1	1980.264.5	False	False	False	2	The American Wing	1980.0	Coin	Ten-dollar Liberty Head Coin	107	...	
2	67.265.9	False	False	False	3	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	NaN	...	
3	67.265.10	False	False	False	4	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	NaN	...	
4	67.265.11	False	False	False	5	The American Wing	1967.0	Coin	Two-and-a-Half Dollar Coin	NaN	...	
...	...	...	...	...	...	...	...	...	...	...	...	
477799	23.112.2893	False	False	True	860869	Drawings and Prints	1923	Drawing	Phaeton #24567	16517	...	Dr
477800	23.112.2894	False	False	True	860870	Drawings and Prints	1923	Drawing	Stanhope Phaeton #25538-25539 (#21222)	16517	...	Dr
477801	53.600.1434	False	False	False	860871	Drawings and Prints	1953	Print	Forest landscape with cattle drinking, a woman...	16489	...	
477802	23.112.2895	False	False	True	860872	Drawings and Prints	1923	Drawing	Phaeton with folding top	16517	...	Dr

	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Department	AccessionYear	Object Name	Title	Constituent ID	...	Classif
477803	23.112.2896	False	False	True	860873	Drawings and Prints	1923	Drawing	Stanhope Phaeton	16517	...	Dr

477804 rows × 53 columns

In [80]: ▶|
```
my_colm = df.columns
my_colm
```

Out[80]: Index(['Object Number', 'Is Highlight', 'Is Timeline Work', 'Is Public Domain',
            'Object ID', 'Gallery Number', 'Department', 'AccessionYear',
            'Object Name', 'Title', 'Culture', 'Period', 'Dynasty', 'Reign',
            'Portfolio', 'Constituent ID', 'Artist Role', 'Artist Prefix',
            'Artist Display Name', 'Artist Display Bio', 'Artist Suffix',
            'Artist Alpha Sort', 'Artist Nationality', 'Artist Begin Date',
            'Artist End Date', 'Artist Gender', 'Artist ULAN URL',
            'Artist Wikidata URL', 'Object Date', 'Object Begin Date',
            'Object End Date', 'Medium', 'Dimensions', 'Credit Line',
            'Geography Type', 'City', 'State', 'County', 'Country', 'Region',
            'Subregion', 'Locale', 'Locus', 'Excavation', 'River', 'Classification',
            'Rights and Reproduction', 'Link Resource', 'Object Wikidata URL',
            'Metadata Date', 'Repository', 'Tags', 'Tags AAT URL',
            'Tags Wikidata URL'],
           dtype='object')

In [69]: ▶| ```python
#Pivot the data

melted = pd.melt(df, id_vars='Department')
melted
```

Out[69]:

|  | Department | variable | value |
|---|---|---|---|
| 0 | The American Wing | Object Number | 1979.486.1 |
| 1 | The American Wing | Object Number | 1980.264.5 |
| 2 | The American Wing | Object Number | 67.265.9 |
| 3 | The American Wing | Object Number | 67.265.10 |
| 4 | The American Wing | Object Number | 67.265.11 |
| ... | ... | ... | ... |
| 25323607 | Drawings and Prints | Tags Wikidata URL | https://www.wikidata.org/wiki/Q235356 |
| 25323608 | Drawings and Prints | Tags Wikidata URL | https://www.wikidata.org/wiki/Q235356 |
| 25323609 | Drawings and Prints | Tags Wikidata URL | NaN |
| 25323610 | Drawings and Prints | Tags Wikidata URL | https://www.wikidata.org/wiki/Q235356 |
| 25323611 | Drawings and Prints | Tags Wikidata URL | https://www.wikidata.org/wiki/Q235356 |

25323612 rows × 3 columns

In [70]:   ▶| 
```python
#using the melt method to go from wide to long format

melted_2 = pd.melt(df, id_vars='Object Name')
melted_2
```

Out[70]:

	Object Name	variable	value
0	Coin	Object Number	1979.486.1
1	Coin	Object Number	1980.264.5
2	Coin	Object Number	67.265.9
3	Coin	Object Number	67.265.10
4	Coin	Object Number	67.265.11
...	...	...	...
25323607	Drawing	Tags Wikidata URL	https://www.wikidata.org/wiki/Q235356
25323608	Drawing	Tags Wikidata URL	https://www.wikidata.org/wiki/Q235356
25323609	Print	Tags Wikidata URL	NaN
25323610	Drawing	Tags Wikidata URL	https://www.wikidata.org/wiki/Q235356
25323611	Drawing	Tags Wikidata URL	https://www.wikidata.org/wiki/Q235356

25323612 rows × 3 columns

In [75]: ▶|
```python
pd.melt(df, value_vars=['Object Name', 'Object ID', 'Title'])
```

Out[75]:

	variable	value
0	Object Name	Coin
1	Object Name	Coin
2	Object Name	Coin
3	Object Name	Coin
4	Object Name	Coin
...	...	...
1433407	Title	Phaeton #24567
1433408	Title	Stanhope Phaeton #25538-25539 (#21222)
1433409	Title	Forest landscape with cattle drinking, a woman...
1433410	Title	Phaeton with folding top
1433411	Title	Stanhope Phaeton

1433412 rows × 2 columns

In [ ]: ▶|
```python
#Chapter 10
```

In [84]: ▶|
```python
df.groupby(by='AccessionYear')
```

Out[84]: <pandas.core.groupby.generic.DataFrameGroupBy object at 0x0000025F12D224C0>

In [88]:   ▶| ` df.groupby(['AccessionYear', 'Object Number']).mean() `

Out[88]:

AccessionYear	Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Object Begin Date	Object End Date	Metadata Date
1870.0	70.1	1.0	1.0	1.0	239584.0	200.0	225.0	NaN
1871.0	71.1	0.0	0.0	1.0	436048.0	1605.0	1669.0	NaN
	71.100	0.0	0.0	1.0	436523.0	1470.0	1499.0	NaN
	71.105	0.0	0.0	1.0	436119.0	1755.0	1805.0	NaN
	71.108	0.0	0.0	1.0	436615.0	1628.0	1628.0	NaN
...	...	...	...	...	...	...	...	...
2022	2022.3	0.0	0.0	0.0	856381.0	1769.0	1769.0	NaN
	2022.4.1–.6	0.0	0.0	0.0	856382.0	1605.0	1615.0	NaN
	2022.5	0.0	0.0	0.0	852556.0	2011.0	2021.0	NaN
	2022.6	0.0	0.0	0.0	854846.0	1495.0	1505.0	NaN
	2022.8	0.0	0.0	0.0	856672.0	1896.0	1896.0	NaN

471626 rows × 7 columns

In [90]: ▶| `df.groupby('Object Number').mean()`

Out[90]:

Object Number	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Object Begin Date	Object End Date	Metadata Date
00.1.1	0.0	0.0	1.0	212639.0	1800.0	1899.0	NaN
00.1.10	0.0	0.0	1.0	212646.0	1800.0	1899.0	NaN
00.1.11	0.0	0.0	1.0	78933.0	1800.0	1899.0	NaN
00.1.14	0.0	0.0	1.0	212649.0	1845.0	1855.0	NaN
00.1.15	0.0	0.0	1.0	212650.0	1800.0	1899.0	NaN
...	...	...	...	...	...	...	...
x.685.9	0.0	0.0	0.0	359743.0	1805.0	1805.0	NaN
x.686.1–.4	0.0	0.0	0.0	359409.0	1762.0	1762.0	NaN
x.7	0.0	0.0	1.0	342932.0	1514.0	1600.0	NaN
x.791	0.0	0.0	1.0	370824.0	1700.0	1900.0	NaN
x.815	0.0	0.0	1.0	388075.0	1600.0	1700.0	NaN

474872 rows × 7 columns

In [91]: ▶| `df.groupby('AccessionYear').mean()`

Out[91]:

	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Object Begin Date	Object End Date	Metadata Date
AccessionYear							
1870.0	1.000000	1.000000	1.000000	239584.000000	200.000000	225.000000	NaN
1871.0	0.000000	0.200000	1.000000	436966.483333	1648.333333	1663.133333	NaN
1872.0	0.250000	0.500000	1.000000	210995.250000	1783.000000	1787.250000	NaN
1873.0	0.000000	0.015152	0.833333	204510.151515	1747.924242	1838.803030	NaN
1874.0	0.001113	0.011869	0.858494	243977.464763	-1622.030972	-290.809347	NaN
...	...	...	...	...	...	...	...
2019	0.014815	0.000000	0.022222	375588.688889	1957.162963	1960.814815	NaN
2020	0.000000	0.000000	0.022556	362115.481203	1958.639098	1959.390977	NaN
2020-03-23	0.000000	0.000000	1.000000	667330.000000	1400.000000	1573.000000	NaN
2021	0.000000	0.000000	0.203840	784308.048744	1849.319055	1864.446086	NaN
2022	0.000000	0.000000	0.000000	855765.714286	1796.571429	1810.285714	NaN

280 rows × 7 columns

In [98]: ▶|
```python
#mapping a dict for grouping. Used Object Number Acession Year and Object ID

mapping = {'Object Number': 'red', 'AccessionYear': 'blue', 'Object ID': 'yellow'}
mapping
```

Out[98]: `{'Object Number': 'red', 'AccessionYear': 'blue', 'Object ID': 'yellow'}`

In [100]:

```
#sum of the dict.
#group by dict/series

by_column = df.groupby(mapping, axis='columns')
by_column.sum()
```

Out[100]:

	blue	red	yellow
0	1979.0	1979.486.1	1
1	1980.0	1980.264.5	2
2	1967.0	67.265.9	3
3	1967.0	67.265.10	4
4	1967.0	67.265.11	5
...	...	...	...
477799	1923	23.112.2893	860869
477800	1923	23.112.2894	860870
477801	1953	53.600.1434	860871
477802	1923	23.112.2895	860872
477803	1923	23.112.2896	860873

477804 rows × 3 columns

In [107]: ▶| *#grouping with a function*

`df.groupby(int).sum()`

Out[107]:

	Is Highlight	Is Timeline Work	Is Public Domain	Object ID	Object Begin Date	Object End Date	Metadata Date
0	0	0	0	1	1853	1853	0.0
1	0	0	0	2	1901	1901	0.0
2	0	0	0	3	1909	1927	0.0
3	0	0	0	4	1909	1927	0.0
4	0	0	0	5	1909	1927	0.0
...	...	...	...	...	...	...	...
477799	0	0	1	860869	1904	1904	0.0
477800	0	0	1	860870	1904	1904	0.0
477801	0	0	0	860871	1768	1778	0.0
477802	0	0	1	860872	1890	1900	0.0
477803	0	0	1	860873	1890	1900	0.0

477804 rows × 7 columns

In [113]: ▶| #selecting the column Acession Year

print(df['AccessionYear'])

```
0          1979.0
1          1980.0
2          1967.0
3          1967.0
4          1967.0
            ...
477799      1923
477800      1923
477801      1953
477802      1923
477803      1923
Name: AccessionYear, Length: 477804, dtype: object
```

In [121]: ▶|
```python
#generate date range from 1979 to 1923

index = pd.date_range('1979', '1923', periods=20)
index
```

Out[121]: DatetimeIndex(['1979-01-01 00:00:00',
                       '1976-01-20 11:22:06.315789472',
                       '1973-02-07 22:44:12.631578944',
                       '1970-02-27 10:06:18.947368416',
                       '1967-03-18 21:28:25.263157888',
                       '1964-04-06 08:50:31.578947328',
                       '1961-04-25 20:12:37.894736832',
                       '1958-05-15 07:34:44.210526336',
                       '1955-06-03 18:56:50.526315776',
                       '1952-06-22 06:18:56.842105216',
                       '1949-07-11 17:41:03.157894656',
                       '1946-07-31 05:03:09.473684224',
                       '1943-08-19 16:25:15.789473664',
                       '1940-09-07 03:47:22.105263104',
                       '1937-09-26 15:09:28.421052672',
                       '1934-10-16 02:31:34.736841984',
                       '1931-11-04 13:53:41.052631552',
                       '1928-11-23 01:15:47.368421120',
                       '1925-12-12 12:37:53.684210432',
                                '1923-01-01 00:00:00'],
                      dtype='datetime64[ns]', freq=None)

In [126]: ▶|
```python
#generate date range used periods of 20 for smaller outputs

pd.date_range(end='1973', periods=20)
```

Out[126]: DatetimeIndex(['1972-12-13', '1972-12-14', '1972-12-15', '1972-12-16',
                       '1972-12-17', '1972-12-18', '1972-12-19', '1972-12-20',
                       '1972-12-21', '1972-12-22', '1972-12-23', '1972-12-24',
                       '1972-12-25', '1972-12-26', '1972-12-27', '1972-12-28',
                       '1972-12-29', '1972-12-30', '1972-12-31', '1973-01-01'],
                      dtype='datetime64[ns]', freq='D')

In [128]: ▶| *#converting to timestamps to periods and back*

```python
dates = pd.date_range('1973', periods=5, freq="M")
```

In [132]: ▶| *#converting to timestamps to periods and back*

```python
ts = pd.Series(np.random.standard_normal(5), index=dates)
ts
```

Out[132]:
```
1973-01-31     0.053080
1973-02-28    -0.452925
1973-03-31     0.342308
1973-04-30    -0.708977
1973-05-31    -0.884298
Freq: M, dtype: float64
```

In [133]: ▶| *#converting to timestamps to periods and back*

```python
pts = ts.to_period()
pts
```

Out[133]:
```
1973-01     0.053080
1973-02    -0.452925
1973-03     0.342308
1973-04    -0.708977
1973-05    -0.884298
Freq: M, dtype: float64
```