Experiment 2

In [1]:
```
import pandas as pd
import numpy as np
```

In [2]:
```
df = pd.read_csv('https://raw.githubusercontent.com/realpython/python-data-cleaning/mast
```

In [3]:
```
df.head()
```

Out[3]:

	Identifier	Edition Statement	Place of Publication	Date of Publication	Publisher	Title	Author	Contributors	Corporat Autho
0	206	NaN	London	1879 [1878]	S. Tinsley & Co.	Walter Forbes. [A novel.] By A. A	A. A.	FORBES, Walter.	Nal
1	216	NaN	London; Virtue & Yorston	1868	Virtue & Co.	All for Greed. [A novel. The dedication signed...	A., A. A.	BLAZE DE BURY, Marie Pauline Rose - Baroness	Nal
2	218	NaN	London	1869	Bradbury, Evans & Co.	Love the Avenger. By the author of “All for Gr...	A., A. A.	BLAZE DE BURY, Marie Pauline Rose - Baroness	Nal
3	472	NaN	London	1851	James Darling	Welsh Sketches, chiefly ecclesiastical, to the...	A., E. S.	Appleyard, Ernest Silvanus.	Nal
4	480	A new edition, revised, etc.	London	1857	Wertheim & Macintosh	[The World in which I live, and my place in it...	A., E. S.	BROOME, John Henry.	Nal

In [4]:
```
len(df.index)
```

Out[4]:
```
8287
```

In [5]:
```
to_drop = ['Edition Statement',
'Corporate Author',
'Corporate Contributors',
'Former owner',
'Engraver',
'Issuance type',
'Flickr URL',
'Shelfmarks']
```

In [6]:
```
df.drop(to_drop, inplace = True, axis = 1)
```

In [7]:
```
df.head()
```

Out[7]:

	Identifier	Place of Publication	Date of Publication	Publisher	Title	Author	Contributors
0	206	London	1879 [1878]	S. Tinsley & Co.	Walter Forbes. [A novel.] By A. A	A. A.	FORBES, Walter.

1	216	London; Virtue & Yorston	1868	Virtue & Co.	All for Greed. [A novel. The dedication signed...	A., A. A.	BLAZE DE BURY, Marie Pauline Rose - Baroness
2	218	London	1869	Bradbury, Evans & Co.	Love the Avenger. By the author of "All for Gr...	A., A. A.	BLAZE DE BURY, Marie Pauline Rose - Baroness
3	472	London	1851	James Darling	Welsh Sketches, chiefly ecclesiastical, to the...	A., E. S.	Appleyard, Ernest Silvanus.
4	480	London	1857	Wertheim & Macintosh	[The World in which I live, and my place in it...	A., E. S.	BROOME, John Henry.

In [8]:
```
df['Identifier'].is_unique
```

Out[8]:
```
True
```

In [9]:
```
df = df.set_index('Identifier')
```

In [10]:
```
df.head()
```

Out[10]:

Identifier	Place of Publication	Date of Publication	Publisher	Title	Author	Contributors
206	London	1879 [1878]	S. Tinsley & Co.	Walter Forbes. [A novel.] By A. A	A. A.	FORBES, Walter.
216	London; Virtue & Yorston	1868	Virtue & Co.	All for Greed. [A novel. The dedication signed...	A., A. A.	BLAZE DE BURY, Marie Pauline Rose - Baroness
218	London	1869	Bradbury, Evans & Co.	Love the Avenger. By the author of "All for Gr...	A., A. A.	BLAZE DE BURY, Marie Pauline Rose - Baroness
472	London	1851	James Darling	Welsh Sketches, chiefly ecclesiastical, to the...	A., E. S.	Appleyard, Ernest Silvanus.
480	London	1857	Wertheim & Macintosh	[The World in which I live, and my place in it...	A., E. S.	BROOME, John Henry.

In [11]:
```
df.loc[216]
```

Out[11]:
```
Place of Publication                            London; Virtue & Yorston
Date of Publication                                                 1868
Publisher                                                  Virtue & Co.
Title                  All for Greed. [A novel. The dedication signed...
Author                                                       A., A. A.
Contributors                BLAZE DE BURY, Marie Pauline Rose - Baroness
Name: 216, dtype: object
```

In [12]:
```
df.iloc[1]
```

Out[12]:
```
Place of Publication                            London; Virtue & Yorston
Date of Publication                                                 1868
Publisher                                                  Virtue & Co.
Title                  All for Greed. [A novel. The dedication signed...
Author                                                       A., A. A.
Contributors                BLAZE DE BURY, Marie Pauline Rose - Baroness
Name: 216, dtype: object
```

In [13]:
```
df.dtypes.value_counts()
```

Out[13]:
```
object    6
dtype: int64
```

In [14]:
```
df.loc[1905:, 'Date of Publication'].head(10)
```

Out[14]:
```
Identifier
1905           1888
1929    1839, 38-54
2836           1897
2854           1865
2956        1860-63
2957           1873
3017           1866
3131           1899
4598           1814
4884           1820
Name: Date of Publication, dtype: object
```

In [15]:
```
extr = df['Date of Publication'].str.extract(r'^(\d{4})', expand=False)
```

In [16]:
```
extr.head()
```

Out[16]:
```
Identifier
206    1879
216    1868
218    1869
472    1851
480    1857
Name: Date of Publication, dtype: object
```

In [17]:
```
df['Place of Publication'].head(10)
```

Out[17]:
```
Identifier
206                                  London
216                 London; Virtue & Yorston
218                                  London
472                                  London
480                                  London
481                                  London
519                                  London
667     pp. 40. G. Bryan & Co: Oxford, 1898
874                                 London]
1143                                 London
Name: Place of Publication, dtype: object
```

In [18]:
```
df.loc[4157862]
```

Out[18]:
```
Place of Publication                                  Newcastle-upon-Tyne
Date of Publication                                                  1867
Publisher                                                     T. Fordyce
Title                  Local Records; or, Historical Register of rema...
Author                     FORDYCE, T. - Printer, of Newcastle-upon-Tyne
Contributors            SYKES, John - Bookseller, of Newcastle-upon-Tyne
Name: 4157862, dtype: object
```

In [19]:
```
df.loc[4159587]
```

Out[19]:
```
Place of Publication                                  Newcastle upon Tyne
Date of Publication                                                  1834
Publisher                                                Mackenzie & Dent
Title                  An historical, topographical and descriptive v...
Author                                                Mackenzie, E. (Eneas)
Contributors                                           ROSS, M. - of Durham
Name: 4159587, dtype: object
```

In [20]:
```
pub = df['Place of Publication']
```

```
london = pub.str.contains('London')
```

In [21]:
```
london
```

Out[21]:
```
Identifier
206          True
216          True
218          True
472          True
480          True
            ...
4158088      True
4158128     False
4159563      True
4159587     False
4160339      True
Name: Place of Publication, Length: 8287, dtype: bool
```

In [22]:
```
oxford = pub.str.contains('Oxford')
```

In [23]:
```
df['Place of Publication'] = np.where(london, 'London', np.where(oxford,'Oxford', pub.st
```

In [24]:
```
df.head(10)
```

Out[24]:

Identifier	Place of Publication	Date of Publication	Publisher	Title	Author	Contributors
206	London	1879 [1878]	S. Tinsley & Co.	Walter Forbes. [A novel.] By A. A	A. A.	FORBES, Walter.
216	London	1868	Virtue & Co.	All for Greed. [A novel. The dedication signed...	A., A. A.	BLAZE DE BURY, Marie Pauline Rose - Baroness
218	London	1869	Bradbury, Evans & Co.	Love the Avenger. By the author of “All for Gr...	A., A. A.	BLAZE DE BURY, Marie Pauline Rose - Baroness
472	London	1851	James Darling	Welsh Sketches, chiefly ecclesiastical, to the...	A., E. S.	Appleyard, Ernest Silvanus.
480	London	1857	Wertheim & Macintosh	[The World in which I live, and my place in it...	A., E. S.	BROOME, John Henry.
481	London	1875	William Macintosh	[The World in which I live, and my place in it...	A., E. S.	BROOME, John Henry.
519	London	1872	The Author	Lagonells. By the author of Darmayne (F. E. A....	A., F. E.	ASHLEY, Florence Emily.
667	Oxford	NaN	NaN	The Coming of Spring, and other poems. By J. A...	A., J.|A., J.	ANDREWS, J. - Writer of Verse
874	London	1676	NaN	A Warning to the inhabitants of England, and L...	Rema‘.	ADAMS, Mary.
1143	London	1679	NaN	A Satyr against Vertue. (A poem: supposed to b...	A., T.	OLDHAM, John.

In [25]:
```
olym = pd.read_csv('https://raw.githubusercontent.com/irJERAD/Intro-to-Data-Science-in-P
```

In [26]:
```
olym.head()
```

Out[26]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

0	NaN	№ Summer	01 !	02 !	03 !	Total	№ Winter	01 !	02 !	03 !	Total	№ Games	01 !	02 !	03 !	Combined total
1	Afghanistan (AFG)	13	0	0	2	2	0	0	0	0	0	13	0	0	2	2
2	Algeria (ALG)	12	5	2	8	15	3	0	0	0	0	15	5	2	8	15
3	Argentina (ARG)	23	18	24	28	70	18	0	0	0	0	41	18	24	28	70
4	Armenia (ARM)	5	1	2	9	12	6	0	0	0	0	11	1	2	9	12

In [27]:
```
olym = pd.read_csv('https://raw.githubusercontent.com/irJERAD/Intro-to-Data-Science-in-P
```

In [28]:
```
olym.head()
```

Out[28]:

	Unnamed: 0	№ Summer	01 !	02 !	03 !	Total	№ Winter	01 !.1	02 !.1	03 !.1	Total.1	№ Games	01 !.2	02 !.2	03 !.2	Comb
0	Afghanistan (AFG)	13	0	0	2	2	0	0	0	0	0	13	0	0	2	
1	Algeria (ALG)	12	5	2	8	15	3	0	0	0	0	15	5	2	8	
2	Argentina (ARG)	23	18	24	28	70	18	0	0	0	0	41	18	24	28	
3	Armenia (ARM)	5	1	2	9	12	6	0	0	0	0	11	1	2	9	
4	Australasia (ANZ) [ANZ]	2	3	4	5	12	0	0	0	0	0	2	3	4	5	

In [29]:
```
new_names = {'Unnamed: 0': 'Country',
'? Summer': 'Summer Olympics',
'01 !': 'Gold',
'02 !': 'Silver',
'03 !': 'Bronze',
'? Winter': 'Winter Olympics',
'01 !.1': 'Gold.1',
'02 !.1': 'Silver.1',
'03 !.1': 'Bronze.1',
'? Games': 'No. of Games',
'01 !.2': 'Gold.2',
'02 !.2': 'Silver.2',
'03 !.2': 'Bronze.2'}
```

In [30]:
```
olym.rename(columns=new_names, inplace=True)
```

In [31]:
```
olym.head(10)
```

Out[31]:

	Country	№ Summer	Gold	Silver	Bronze	Total	№ Winter	Gold.1	Silver.1	Bronze.1	Total.1	I Game
0	Afghanistan (AFG)	13	0	0	2	2	0	0	0	0	0	
1	Algeria (ALG)	12	5	2	8	15	3	0	0	0	0	
2	Argentina (ARG)	23	18	24	28	70	18	0	0	0	0	
3	Armenia (ARM)	5	1	2	9	12	6	0	0	0	0	
4	Australasia (ANZ) [ANZ]	2	3	4	5	12	0	0	0	0	0	
5	Australia (AUS) [AUS] [Z]	25	139	152	177	468	18	5	3	4	12	
6	Austria (AUT)	26	18	33	35	86	22	59	78	81	218	
7	Azerbaijan (AZE)	5	6	5	15	26	5	0	0	0	0	

8	Bahamas (BAH)	15	5	2	5	12	0	0	0	0	0
9	Bahrain (BRN)	8	0	0	1	1	0	0	0	0	0

In [ ]: