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]:
                            Edition
                                       Place of
                                                     Date of
                                                                                                               Corporate
                                                              Publisher
                                                                                  Title Author Contributors
             Identifier
                        Statement Publication Publication
                                                                                                                  Autho
                                                                                Walter
                                                               S. Tinsley
                                                                            Forbes. [A
                                                                                                     FORBES,
          0
                   206
                              NaN
                                        London 1879 [1878]
                                                                                          A. A.
                                                                                                                    Nal
                                                                  & Co.
                                                                           novel.] By A.
                                                                                                      Walter.
                                                                          All for Greed.
                                                                                                    BLAZE DE
                                        London;
                                                                Virtue &
                                                                          [A novel. The
                                                                                          A., A.
                                                                                                 BURY, Marie
                   216
                                        Virtue &
                                                       1868
          1
                              NaN
                                                                                                                    Nal
                                                                    Co.
                                                                            dedication
                                                                                            Α.
                                                                                                 Pauline Rose
                                        Yorston
                                                                                                   - Baroness
                                                                              signed...
                                                                                                   BLAZE DE
                                                                              Love the
                                                               Bradbury,
                                                                                         A., A.
                                                                                                 BURY, Marie
                                                                           Avenger. By
          2
                   218
                              NaN
                                        London
                                                       1869
                                                                Evans &
                                                                                                                    Nat
                                                                          the author of
                                                                                            Α.
                                                                                                 Pauline Rose
                                                                    Co.
                                                                            "All for Gr...
                                                                                                   - Baroness
                                                                                Welsh
                                                                             Sketches,
                                                                                                   Appleyard,
                                                                 James
                                                                                          A., E.
          3
                   472
                              NaN
                                        London
                                                        1851
                                                                                chiefly
                                                                                                       Ernest
                                                                                                                    Nal
                                                                 Darling
                                                                                            S.
                                                                         ecclesiastical,
                                                                                                     Silvanus.
                                                                               to the...
                             A new
                                                                          [The World in
                                                               Wertheim
                            edition,
                                                                           which I live,
                                                                                          A., E.
                                                                                                    BROOME,
                   480
                                                        1857
          4
                                        London
                                                                     &
                                                                                                                    Nal
                           revised,
                                                                          and my place
                                                                                            S.
                                                                                                  John Henry.
                                                              Macintosh
                               etc.
                                                                                 in it...
In [4]:
          len(df.index)
Out[4]:
In [5]:
          to drop = ['Edition Statement',
          'Corporate Author',
           'Corporate Contributors',
           'Former owner',
           'Engraver',
           'Issuance type',
           'Flickr URL',
           'Shelfmarks']
          df.drop(to drop, inplace = True, axis = 1)
In [6]:
          df.head()
In [7]:
Out[7]:
                            Place of
                                          Date of
             Identifier
                                                    Publisher
                                                                                 Title Author
                                                                                                       Contributors
                         Publication
                                     Publication
                                                     S. Tinsley
                                                               Walter Forbes. [A novel.]
          0
                   206
                             London 1879 [1878]
                                                                                                     FORBES, Walter.
                                                                                          A. A.
                                                        & Co.
                                                                               By A. A
```

	1 2°	16 Lond Virtue Yors	e &	68 Virtue Co	-								
	2 2°			Bradbury 69 Evans Co	& author of "All for G		Marie Palline Rose -						
	3 47			Jame Darlin	•								
	4 48	30 Lond	don 18	Wertheir 57 Macintos	& live and my place in it								
In [8]:	df['Identifier'].is_unique												
Out[8]:	True												
In [9]:	<pre>df = df.set_index('Identifier')</pre>												
n [10]:	df.head()												
out[10]:		Place of Publication		Publisher	Title	e Author	Contributors						
	Identifier	Publication	rubilcation										
	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		BLAZE DE BURY, Marie Pauline Rose - Baroness						
	218	London	1869	Bradbury, Evans & Co.	Love the Avenger. By the author of "All for Gr		BLAZE DE BURY, Marie Pauline Rose - Baroness						
	472	London	1851	James Darling	Welsh Sketches, chiefly ecclesiastical, to the		Appleyard, Ernest Silvanus.						
	480	London	1857	Wertheim & Macintosh	[The World in which I live	, A., E.	BROOME, John Henry.						
				Macintosii	and my place in it		BroomE, committently.						
n [11]:	df.loc[2]	16]		Macintosii	and my place in it		BROOME, COMMITTERNY.						
	Place of Date of Publisher Title Author Contribut	Publicatio Publication r	All fo	or Greed. [/	London; Vi: A novel. The dedicate RY, Marie Pauline Ro	rtue & Yo Virtue tion sign	orston 1868 & Co. ned A. A.						
n [11]: ut[11]: n [12]:	Place of Date of Publisher Title Author Contribut	Publication r tors 6, dtype: o	All fo	or Greed. [/	London; Vin	rtue & Yo Virtue tion sign	orston 1868 & Co. ned A. A.						
ut[11]:	Place of Publisher Title Author Contribut Name: 216 Place of Date of Publisher Title Author Contribut Contribut Contribut Contribut Contribut Contribut Contribut	Publication r tors 6, dtype: o 1] Publication Publication r	All for All fo	or Greed. [A	London; Vin	Virtue tion sign A., ose - Ban rtue & Yo Virtue tion sign A.,	orston 1868 & Co. ned A. A. coness orston 1868 & Co. ned A. A.						

```
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
                       1866
         3017
         3131
                       1899
                        1814
         4598
         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()
         Identifier
Out[16]:
         206 1879
         216
                1868
         218
                1869
             1851
         472
         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
               pp. 40. G. Bryan & Co: Oxford, 1898
         667
         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
         Publisher
                                                                       T. Fordyce
         Title
                                Local Records; or, Historical Register of rema...
                                  FORDYCE, T. - Printer, of Newcastle-upon-Tyne
         Author
                                  SYKES, John - Bookseller, of Newcastle-upon-Tyne
         Contributors
         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
In [21]:
            Identifier
Out[21]:
            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
            oxford = pub.str.contains('Oxford')
In [22]:
In [23]:
            df['Place of Publication'] = np.where(london, 'London', np.where(oxford,'Oxford', pub.st
            df.head(10)
In [24]:
Out [24]:
                          Place of
                                        Date of
                                                  Publisher
                                                                                  Title Author
                                                                                                           Contributors
                       Publication
                                    Publication
            Identifier
                                                                Walter Forbes. [A novel.]
                                                 S. Tinsley &
                 206
                           London
                                    1879 [1878]
                                                                                           A. A.
                                                                                                        FORBES, Walter.
                                                                                By A. A
                                                         Co.
                                                                                                  BLAZE DE BURY, Marie
                                                    Virtue &
                                                              All for Greed. [A novel. The
                                                                                           A., A.
                 216
                           London
                                          1868
                                                                                                          Pauline Rose -
                                                                     dedication signed...
                                                         Co.
                                                                                             Α.
                                                                                                               Baroness
                                                   Bradbury,
                                                                                                  BLAZE DE BURY, Marie
                                                                Love the Avenger. By the
                                                                                           A., A.
                 218
                                                    Evans &
                                                                                                          Pauline Rose -
                           London
                                          1869
                                                                   author of "All for Gr...
                                                                                             Α.
                                                                                                               Baroness
                                                         Co.
                                                                                                       Appleyard, Ernest
                                                      James
                                                                 Welsh Sketches, chiefly
                                                                                           A., E.
                 472
                           London
                                           1851
                                                     Darling
                                                                  ecclesiastical, to the...
                                                                                             S.
                                                                                                               Silvanus.
                                                 Wertheim &
                                                               [The World in which I live,
                                                                                           A., E.
                 480
                           London
                                          1857
                                                                                                   BROOME, John Henry.
                                                  Macintosh
                                                                     and my place in it...
                                                                                             S.
                                                     William
                                                               [The World in which I live,
                                                                                           A., E.
                 481
                           London
                                          1875
                                                                                                   BROOME, John Henry.
                                                  Macintosh
                                                                     and my place in it...
                                                              Lagonells. By the author of
                                                                                           A., F.
                                                                                                       ASHLEY, Florence
                                                 The Author
                 519
                           London
                                          1872
                                                                    Darmayne (F. E. A....
                                                                                              Ε.
                                                                                                                  Emily.
                                                              The Coming of Spring, and
                                                                                                   ANDREWS, J. - Writer
                                                                                             Α.,
                 667
                           Oxford
                                           NaN
                                                        NaN
                                                                  other poems. By J. A...
                                                                                         J.|A., J.
                                                                                                                of Verse
                                                                        A Warning to the
                 874
                           London
                                          1676
                                                        NaN
                                                                 inhabitants of England,
                                                                                         Rema<sup>c</sup>.
                                                                                                          ADAMS, Mary.
                                                                                and L...
                                                               A Satyr against Vertue. (A
                1143
                           London
                                          1679
                                                        NaN
                                                                                                         OLDHAM, John.
                                                                                           A., T.
                                                                 poem: supposed to b...
In [25]:
            olym = pd.read csv('https://raw.githubusercontent.com/irJERAD/Intro-to-Data-Science-in-P
In [26]:
            olym.head()
                               0
                                        1
                                             2
                                                 3
                                                     4
                                                            5
                                                                        7
                                                                                 9
                                                                                       10
                                                                                                11 12 13 14
                                                                                                                        15
Out[26]:
                                                                    6
                                                                             8
```

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

```
Summer
                                                            Winter
                                                                                       Games
                                                   İ
                                                                              !
                                                                                                    Ţ
                                                                                                                total
                                                                         !
                                                                                                        !
                                                   2
            1 Afghanistan (AFG)
                                     13
                                               0
                                                                 0
                                                                     0
                                                                         0
                                                                             0
                                                                                           13
                                                                                                0
                                                                                                    0
                                                                                                        2
                                                                                                                   2
           2
                                               2
                                                                         0
                                                                                                5
                                                                                                    2
                                                                                                                  15
                   Algeria (ALG)
                                      12
                                           5
                                                   8
                                                        15
                                                                 3
                                                                     0
                                                                             0
                                                                                    0
                                                                                           15
                                                                                                        8
           3
                Argentina (ARG)
                                     23
                                        18
                                             24
                                                  28
                                                        70
                                                                18
                                                                     0
                                                                         0
                                                                             0
                                                                                               18
                                                                                                   24
                                                                                                       28
                                                                                                                  70
                                                                                           41
                                               2
                                                                         0
                                                                             0
                                                                                    0
           4
                 Armenia (ARM)
                                      5
                                           1
                                                   9
                                                        12
                                                                 6
                                                                     0
                                                                                           11
                                                                                                    2
                                                                                                        9
                                                                                                                  12
           olym = pd.read csv('https://raw.githubusercontent.com/irJERAD/Intro-to-Data-Science-in-P
In [27]:
           olym.head()
In [28]:
                                                                  Nº
Out[28]:
                                      Nº
                                          01
                                              02
                                                   03
                                                                      01 02
                                                                              03
                                                                                               Nº
                                                                                                   01
                                                                                                        02
                                                                                                            03
                                                                                                                 Comb
                                                       Total
                                                                                   Total.1
                   Unnamed: 0
                                Summer
                                                             Winter
                                                                                           Games
                                                ļ
                                                                     !.1
                                                                          !.1
                                                                              !.1
                                                                                                   1.2
                                                                                                        1.2
                                                                                                            !.2
                                                    2
                                                           2
                                                                           0
                                                                                0
                                                                                        0
                                                                                                         0
                                                                                                              2
           O Afghanistan (AFG)
                                      13
                                           0
                                               0
                                                                   0
                                                                       0
                                                                                               13
                                                                                                     0
            1
                   Algeria (ALG)
                                      12
                                           5
                                                2
                                                    8
                                                          15
                                                                   3
                                                                       0
                                                                           0
                                                                                0
                                                                                        0
                                                                                               15
                                                                                                     5
                                                                                                         2
                                                                                                              8
           2
                                              24
                                                                  18
                                                                       0
                                                                           0
                                                                                        0
                                                                                                   18
                                                                                                        24
                                                                                                            28
                Argentina (ARG)
                                      23
                                          18
                                                   28
                                                          70
                                                                                0
                                                                                               41
           3
                 Armenia (ARM)
                                                2
                                                    9
                                                          12
                                                                   6
                                                                       0
                                                                           0
                                                                                0
                                                                                        0
                                                                                                11
                                                                                                     1
                                                                                                         2
                                                                                                              9
               Australasia (ANZ)
                                       2
                                           3
                                                                   0
                                                                       0
                                                                           0
                                                                                        0
                                                                                                2
                                                                                                     3
                                                                                                              5
                                               4
                                                    5
                                                          12
                                                                                0
                                                                                                         4
                         [ANZ]
           new_names = {'Unnamed: 0': 'Country',
In [29]:
           '? 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'}
           olym.rename(columns=new names, inplace=True)
In [30]:
           olym.head(10)
In [31]:
                                      Nº
Out[31]:
                                                                           Nº
                                          Gold Silver Bronze Total
                                                                               Gold.1 Silver.1 Bronze.1 Total.1
                       Country
                                                                       Winter
                                Summer
           O Afghanistan (AFG)
                                             0
                                                             2
                                                                    2
                                                                                                      0
                                                                                                              0
                                      13
                                                    0
                                                                            0
                                                                                   0
                                                                                            0
           1
                   Algeria (ALG)
                                      12
                                             5
                                                    2
                                                             8
                                                                   15
                                                                            3
                                                                                    0
                                                                                            0
                                                                                                              0
           2
                                                                   70
                                                                                            0
                                                                                                              0
                Argentina (ARG)
                                      23
                                             18
                                                   24
                                                            28
                                                                           18
                                                                                    0
                                                     2
           3
                 Armenia (ARM)
                                       5
                                                             9
                                                                   12
                                                                            6
                                                                                            0
                                                                                                              0
               Australasia (ANZ)
                                       2
                                                                                                              0
           4
                                             3
                                                             5
                                                                   12
                                                                            0
                                                                                   0
                                                                                            0
                                                                                                      0
                                                    4
                         [ANZ]
                 Australia (AUS)
                                                                 468
           5
                                      25
                                           139
                                                  152
                                                           177
                                                                           18
                                                                                    5
                                                                                            3
                                                                                                      4
                                                                                                             12
                      [AUS] [Z]
                                      26
           6
                   Austria (AUT)
                                            18
                                                   33
                                                            35
                                                                  86
                                                                           22
                                                                                  59
                                                                                           78
                                                                                                     81
                                                                                                            218
                Azerbaijan (AZE)
                                       5
                                             6
                                                            15
                                                                   26
                                                                            5
```

NaN

Nº 01 02 03 Total

Nº 01 02 03 Total

Nº 01 02 03 Combined

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 []: