numpy and pandas

September 11, 2017

```
In [46]: import numpy as np
         import pandas as pd
```

1 Numpy

```
In [13]: a1 = np.zeros(10)
         a1

Out[13]: array([ 0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.])

In [14]: a1[5] = 1
         a1

Out[14]: array([ 0.,  0.,  0.,  0.,  0.,  1.,  0.,  0.,  0.,  0.])

In [15]: a2 = np.array(range(10, 50))
         a2

Out[15]: array([10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26,
                27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43,
                44, 45, 46, 47, 48, 49])

In [16]: a3 = np.array(range(25))
         new_a3 = np.reshape(a3,(5,5))
         mat1 = np.asmatrix(new_a3)
         mat1

Out[16]: matrix([[ 0,  1,  2,  3,  4],
                 [ 5,  6,  7,  8,  9],
                 [10, 11, 12, 13, 14],
                 [15, 16, 17, 18, 19],
                 [20, 21, 22, 23, 24]])

In [20]: a4 = np.eye(5,5)
         mat2 = np.asmatrix(a4)
         mat2
```

```
Out[20]: matrix([[ 1.,  0.,  0.,  0.,  0.],
                 [ 0.,  1.,  0.,  0.,  0.],
                 [ 0.,  0.,  1.,  0.,  0.],
                 [ 0.,  0.,  0.,  1.,  0.],
                 [ 0.,  0.,  0.,  0.,  1.]])

In [24]: a5 = np.random.rand(5,5)

In [25]: a5.min(), a5.max()

Out[25]: (0.044105603811281591, 0.99281417043248632)

In [29]: a6 = np.ones((4,3))
         a7 = np.random.rand(3,2)
         result = a6.dot(a7)
         result

Out[29]: array([[ 1.94984428,  1.30937809],
                [ 1.94984428,  1.30937809],
                [ 1.94984428,  1.30937809],
                [ 1.94984428,  1.30937809]])

In [33]: np.transpose(result)

Out[33]: array([[ 1.94984428,  1.94984428,  1.94984428,  1.94984428],
                [ 1.30937809,  1.30937809,  1.30937809,  1.30937809]])

In [40]: a8 = np.array(range(25))
         new_a8 = np.reshape(a8, (5,5))

         a9 = np.array(range(25, 50))
         new_a9 = np.reshape(a9, (5,5))

In [43]: add = new_a8 + new_a9
         sub = new_a8 - new_a9
         add

Out[43]: array([[25, 27, 29, 31, 33],
                [35, 37, 39, 41, 43],
                [45, 47, 49, 51, 53],
                [55, 57, 59, 61, 63],
                [65, 67, 69, 71, 73]])

In [45]: sub

Out[45]: array([[-25, -25, -25, -25, -25],
                [-25, -25, -25, -25, -25],
                [-25, -25, -25, -25, -25],
                [-25, -25, -25, -25, -25],
                [-25, -25, -25, -25, -25]])
```

2 Pandas

```
In [48]: exam_data = {"name": ["Anastasia", "Catherine", "Cahill", "James", "Emily", "Michael",
                     'score':[13,9.5,16.5,np.nan,11,20,17,np.nan,8.5,19],
                     'attempts':[1,3,3,2,2,3,2,3,2,1],
                     'qualify':['yes','no','yes','no','no','yes','yes','no','no','yes']}
         labels = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']
         df = pd.DataFrame(exam_data, index = labels)
         df
```

```
Out[48]:    attempts       name qualify  score
         a         1  Anastasia     yes   13.0
         b         3  Catherine      no    9.5
         c         3     Cahill     yes   16.5
         d         2      James      no    NaN
         e         2      Emily      no   11.0
         f         3    Michael     yes   20.0
         g         2     Monica     yes   17.0
         h         3      Laura      no    NaN
         i         2      Kevin      no    8.5
         j         1     Jordan     yes   19.0
```

```
In [49]: df.loc[:, ['name', 'score']]
```

```
Out[49]:         name  score
         a  Anastasia   13.0
         b  Catherine    9.5
         c     Cahill   16.5
         d      James    NaN
         e      Emily   11.0
         f    Michael   20.0
         g     Monica   17.0
         h      Laura    NaN
         i      Kevin    8.5
         j     Jordan   19.0
```

```
In [50]: df.iloc[:3, :]
```

```
Out[50]:    attempts       name qualify  score
         a         1  Anastasia     yes   13.0
         b         3  Catherine      no    9.5
         c         3     Cahill     yes   16.5
```

```
In [51]: df.iloc[[1,2,5,6], [1,3]]
```

```
Out[51]:         name  score
         b  Catherine    9.5
         c     Cahill   16.5
         f    Michael   20.0
         g     Monica   17.0
```

```
In [52]: df[df['attempts'] > 2]

Out[52]:    attempts       name qualify  score
         b         3  Catherine      no    9.5
         c         3     Cahill     yes   16.5
         f         3    Michael     yes   20.0
         h         3      Laura      no    NaN

In [53]: df[df['score'].isnull()]

Out[53]:    attempts   name qualify  score
         d         2  James      no    NaN
         h         3  Laura      no    NaN

In [55]: df[(df['attempts']>2) & (df['score']<15)]

Out[55]:    attempts       name qualify  score
         b         3  Catherine      no    9.5

In [56]: df['attempts'].sum()

Out[56]: 22

In [57]: df['score'].mean()

Out[57]: 14.3125

In [60]: df.loc['k', :] = {'name':'Saya', 'score':17.5, 'attempts':2, 'qualify':'yes'}
         df

Out[60]:   attempts       name qualify score
         a        1  Anastasia     yes    13
         b        3  Catherine      no   9.5
         c        3     Cahill     yes  16.5
         d        2      James      no   NaN
         e        2      Emily      no    11
         f        3    Michael     yes    20
         g        2     Monica     yes    17
         h        3      Laura      no   NaN
         i        2      Kevin      no   8.5
         j        1     Jordan     yes    19
         k        2       Saya     yes  17.5

In [65]: df = df.drop('k')
         df

Out[65]:   attempts       name qualify score
         a        1  Anastasia     yes    13
         b        3  Catherine      no   9.5
         c        3     Cahill     yes  16.5
```

```
        d         2      James      no   NaN
        e         2      Emily      no    11
        f         3    Michael     yes    20
        g         2     Monica     yes    17
        h         3      Laura      no   NaN
        i         2      Kevin      no   8.5
        j         1     Jordan     yes    19
```

```
In [67]: df.drop('attempts', axis = 1)
```

```
Out[67]:   attempts        name qualify score
        a         1  Anastasia     yes    13
        b         3  Catherine      no   9.5
        c         3     Cahill     yes  16.5
        d         2      James      no   NaN
        e         2      Emily      no    11
        f         3    Michael     yes    20
        g         2     Monica     yes    17
        h         3      Laura      no   NaN
        i         2      Kevin      no   8.5
        j         1     Jordan     yes    19
```

```
In [72]: df.groupby('attempts')['score'].sum()
```

```
Out[72]: attempts
         1.0    32.0
         2.0    36.5
         3.0    46.0
         Name: score, dtype: float64
```

```
In [74]: exam2_data = {"name": ["Anastasia", "Catherine", "Ronaldo", "James", "Messi", "Michael"
                     'score':[11,20,16.5,np.nan,10,15,20,np.nan,8,8]}
         labels2 = ['a', 'b', 'c', 'd', 'e', 'f', 'g', 'h', 'i', 'j']
         df2 = pd.DataFrame(exam2_data, index = labels2)
         df2
```

```
Out[74]:         name  score
         a  Anastasia   11.0
         b  Catherine   20.0
         c    Ronaldo   16.5
         d      James    NaN
         e      Messi   10.0
         f    Michael   15.0
         g     Monica   20.0
         h      Laura    NaN
         i    Klassen    8.0
         j      Jonas    8.0
```

```
In [76]: new_df = pd.merge(df, df2, how='inner', left_on='name', right_on='name')
         new_df
```

```
Out[76]:   attempts       name qualify score_x  score_y
         0        1  Anastasia     yes      13     11.0
         1        3  Catherine      no     9.5     20.0
         2        2      James      no     NaN      NaN
         3        3    Michael     yes      20     15.0
         4        2     Monica     yes      17     20.0
         5        3      Laura      no     NaN      NaN
```