HW10_Team1

September 13, 2017

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

1 Numpy

```
In [2]: a1 = np.zeros(10)
        a1

Out[2]: array([ 0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.])

In [3]: a1[4] = 1
        a1

Out[3]: array([ 0.,  0.,  0.,  0.,  1.,  0.,  0.,  0.,  0.,  0.])

In [4]: a2 = np.array(range(10, 50))
        a2

Out[4]: 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 [5]: a3 = np.array(range(25))
        new_a3 = np.reshape(a3,(5,5))
        mat1 = np.asmatrix(new_a3)
        mat1

Out[5]: 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 [6]: a4 = np.eye(5,5)
        mat2 = np.asmatrix(a4)
        mat2
```

```
Out[6]: 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 [7]: a5 = np.random.rand(5,5)
```

```
In [8]: a5.min(), a5.max()
```

```
Out[8]: (0.075266121485626947, 0.89199718143094564)
```

```
In [9]: a6 = np.ones((4,3))
        a7 = np.random.rand(3,2)
        result = a6.dot(a7)
        result
```

```
Out[9]: array([[ 2.35035181,  2.40868954],
               [ 2.35035181,  2.40868954],
               [ 2.35035181,  2.40868954],
               [ 2.35035181,  2.40868954]])
```

```
In [10]: np.transpose(result)
```

```
Out[10]: array([[ 2.35035181,  2.35035181,  2.35035181,  2.35035181],
                [ 2.40868954,  2.40868954,  2.40868954,  2.40868954]])
```

```
In [11]: 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 [12]: print('new_a8 \n',new_a8 )
         print('new_a9 \n',new_a9 )
```

```
new_a8
 [[ 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]]
new_a9
 [[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 [13]: add = new_a8 + new_a9
         sub = new_a8 - new_a9
         add

Out[13]: 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 [14]: sub

Out[14]: 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 [15]: 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[15]:    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 [16]: df.loc[:,['name', 'score']]

Out[16]:         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 [17]: df[:3]
```

```
Out[17]:    attempts       name qualify  score
        a         1  Anastasia     yes   13.0
        b         3  Catherine      no    9.5
        c         3     Cahill     yes   16.5
```

```
In [18]: df[['name','score']].iloc[[1,2,5,6]]
```

```
Out[18]:         name  score
        b  Catherine    9.5
        c     Cahill   16.5
        f    Michael   20.0
        g     Monica   17.0
```

```
In [19]: df[df['attempts'] > 2]
```

```
Out[19]:    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 [20]: df[df['score'].isnull()]
```

```
Out[20]:    attempts   name qualify  score
        d         2  James      no    NaN
        h         3  Laura      no    NaN
```

```
In [21]: df[(df['attempts']<2) & (df['score']>15)]
```

```
Out[21]:    attempts    name qualify  score
        j         1  Jordan     yes   19.0
```

```
In [22]: df['attempts'].sum()
```

```
Out[22]: 22
```

```
In [23]: df['score'].mean()
```

```
Out[23]: 14.3125
```

```
In [24]: df
```

```
Out[24]:    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 [25]: df.loc['k'] = {'name':'Saya', 'score':17.5, 'attempts':2, 'qualify':'yes'}
```

```
In [26]: df
```

```
Out[26]:    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
         k         2       Saya     yes   17.5
```

```
In [27]: df = df.drop('k')
         df
```

```
Out[27]:    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 [28]: df.drop('attempts', axis = 1)
```

```
Out[28]:         name qualify  score
         a  Anastasia     yes   13.0
         b  Catherine      no    9.5
         c     Cahill     yes   16.5
         d      James      no    NaN
         e      Emily      no   11.0
         f    Michael     yes   20.0
         g     Monica     yes   17.0
         h      Laura      no    NaN
         i      Kevin      no    8.5
         j     Jordan     yes   19.0
```

```
In [29]: df.groupby('attempts')['score'].sum()
```

```
Out[29]: attempts
         1    32.0
         2    36.5
         3    46.0
         Name: score, dtype: float64
```

```
In [30]: 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[30]:         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 [31]: new_df = pd.merge(df, df2, how='inner', left_on='name', right_on='name')
         new_df
```

```
Out[31]:    attempts       name qualify  score_x  score_y
         0         1  Anastasia     yes     13.0     11.0
         1         3  Catherine      no      9.5     20.0
         2         2      James      no      NaN      NaN
         3         3    Michael     yes     20.0     15.0
         4         2     Monica     yes     17.0     20.0
         5         3      Laura      no      NaN      NaN
```