0.1 Pandas

In [8]:

Series i **DataFrame**

Series

DataFrame

NaN **isnull** i **notnull**

In [2]:

In [9]:

```
<class 'tuple'>
0    (1, 1)
1    (2, 2)
2    (3, 3)
dtype: object
<class 'numpy.ndarray'>
```

In [10]:

```
<bound method Series.keys of 0    (1, 1)
1    (2, 2)
2    (3, 3)
dtype: object>
```

```
In [11]: # ▯▯▯▯
         points.index
         #▯
```

```
Out[11]: RangeIndex(start=0, stop=3, step=1)
```

```
In [12]: # ▯▯▯▯
         points = pd.Series([(1, 1) , (2, 2), (3, 3)], index = ['x', 'y', 'z'])
         points
```

```
Out[12]: x    (1, 1)
         y    (2, 2)
         z    (3, 3)
         dtype: object
```

```
In [13]: points.index
```

```
Out[13]: Index(['x', 'y', 'z'], dtype='object')
```

```
In [41]: # ▯▯▯▯
         points['x']
```

```
Out[41]: (1, 1)
```

```
In [14]: # ▯▯▯▯
         data = {
             'name': ['Anna', 'John', 'Peter', 'Linda'],
             'location' : ['Berlin', 'New York', 'Belgrade', 'London'],
             'age' : [24, 30, 33, 21]
         }
         print(type(data))

         users = pd.DataFrame(data)
```

```
<class 'dict'>
```

```
In [43]: #▯▯▯▯▯
         # ▯▯▯▯▯▯▯▯▯▯▯▯
         users
```

```
Out[43]:    age  location   name
         0   24    Berlin   Anna
         1   30  New York   John
         2   33  Belgrade  Peter
         3   21    London  Linda
```

```
In [44]: # 列名
         users.columns

Out[44]: Index(['age', 'location', 'name'], dtype='object')

In [45]: # 索引
         users.index

Out[45]: RangeIndex(start=0, stop=4, step=1)

In [46]: # 值
         users.v       alues

Out[46]: array([[24, 'Berlin', 'Anna'],
                [30, 'New York', 'John'],
                [33, 'Belgrade', 'Peter'],
                [21, 'London', 'Linda']], dtype=object)

In [47]: # 按列
         print(users['age'])
         print(users.age)

0    24
1    30
2    33
3    21
Name: age, dtype: int64
0    24
1    30
2    33
3    21
Name: age, dtype: int64


In [50]: # 按行
         users.ix[3]

Out[50]: age             21
         location    London
         name         Linda
         Name: 3, dtype: object

In [51]: # 按行切片
         users.ix[1:5]

Out[51]:    age  location   name
         1   30  New York   John
         2   33  Belgrade  Peter
         3   21    London  Linda
```

```
In [52]: # [illegible]
         users.ix[1]['name']

Out[52]: 'John'

In [53]: type(users)

Out[53]: pandas.core.frame.DataFrame

In [54]: type(users['age'])

Out[54]: pandas.core.series.Series

In [55]: type(users.ix[1]['age'])

Out[55]: numpy.int64
```

[illegible]

```
In [63]: # [illegible]
         print(users[users.age\>25])
         print("[illegible] ")
         print(users.age\>25)

   age  location   name
1   30  New York   John
2   33  Belgrade  Peter


0    False
1     True
2     True
3    False
Name: age, dtype: bool
```

[illegible]

```
In [68]: users = users.reindex(columns = ['age', 'location', 'name', 'email'])
         print(users)

         users = users.reindex(index=[0, 1, 'x', 'y', 'z', 'w'])
         print(users)

   age  location   name  email
0   24    Berlin   Anna    NaN
1   30  New York   John    NaN
2   33  Belgrade  Peter    NaN
3   21    London  Linda    NaN
```

```
    age  location  name  email
0  24.0    Berlin  Anna    NaN
1  30.0  New York  John    NaN
x   NaN       NaN   NaN    NaN
y   NaN       NaN   NaN    NaN
z   NaN       NaN   NaN    NaN
w   NaN       NaN   NaN    NaN
```

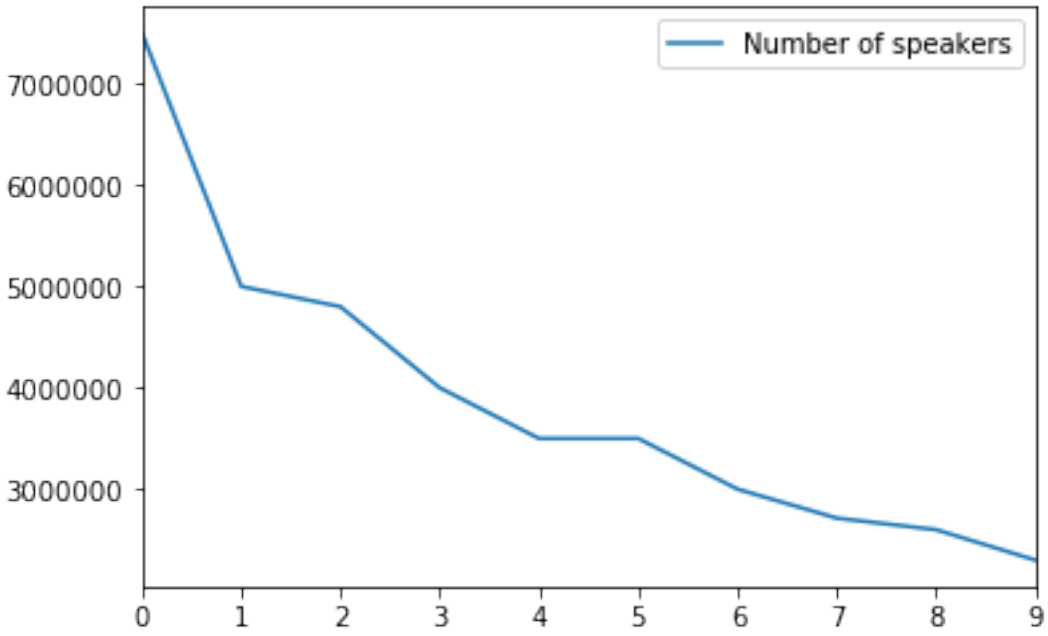
isnull i notnull

```
In [80]: pd.isnull(users['email'])

Out[80]: 0    True
         1    True
         2    True
         3    True
         Name: email, dtype: bool
```

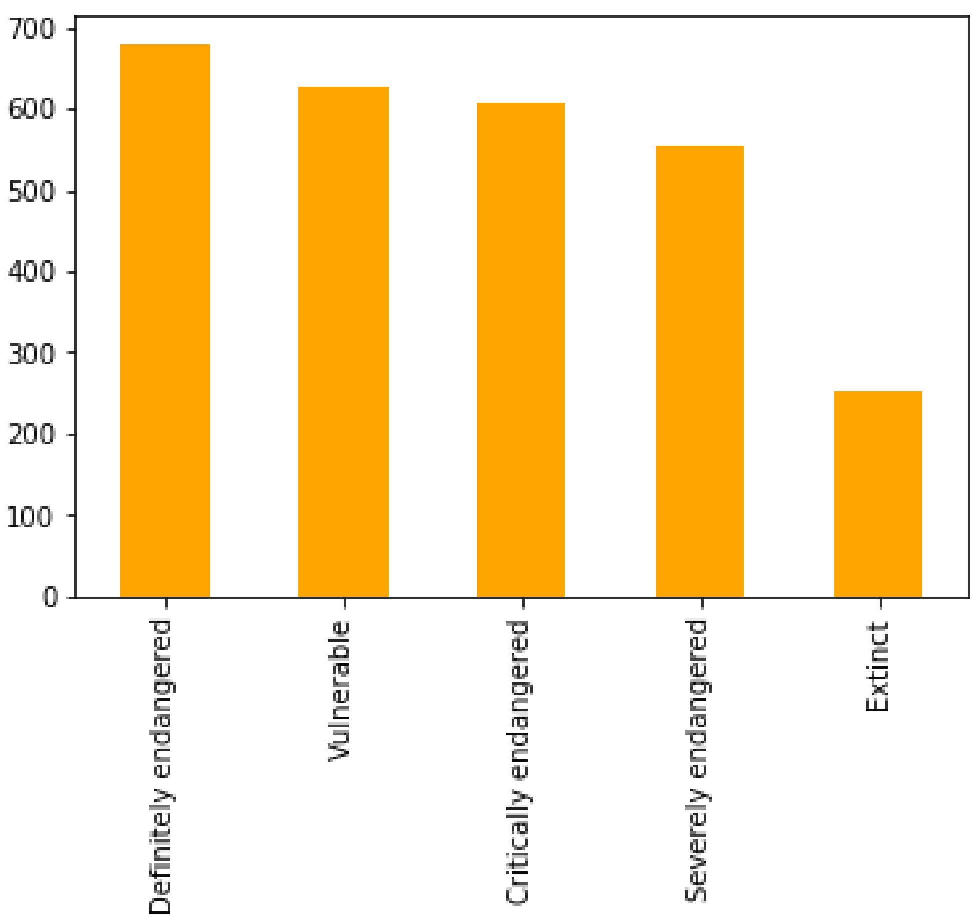
Primer korišćenja raspoloživih podataka

```
In [15]: # [illegible]
         # [illegible]
         # [illegible]
         # [illegible]
         # [illegible]
         # [illegible]
         # [illegible]
         # [illegible]
         # [illegible]

         languages = pd.read_csv        ('data/languages.csv        ')

In [70]: # [illegible]
         # [illegible]

         # [illegible]

In [71]: # [illegible]
         type(languages)

Out[71]: pandas.core.frame.DataFrame

In [72]: # [illegible]
         len(languages)

Out[72]: 2722

In [1]: # [illegible]
        # [illegible]
```

```
In [2]: # [illegible]
        # [illegible]

In [3]: # [illegible]
        # [illegible]

In [76]: # [illegible]
         languages['Name in English'][:14:2]

Out[76]: 0             South Italian
         2                 Low Saxon
         4                   Lombard
         6          Yiddish (Israel)
         8      Limburgian-Ripuarian
         10                  Kumaoni
         12         Emilian-Romagnol
         Name: Name in English, dtype: object

In [4]: # [illegible]
        # [illegible]

In [5]: # [illegible]
        # [illegible]

In [79]: # [illegible]
         [illegible]
         \%[illegible]          inline

         languages[['Name in English', 'Number of speakers']][:10].plot()

Out[79]: <matplotlib.axes._subplots.AxesSubplot at 0xc78b21f4e0>
```

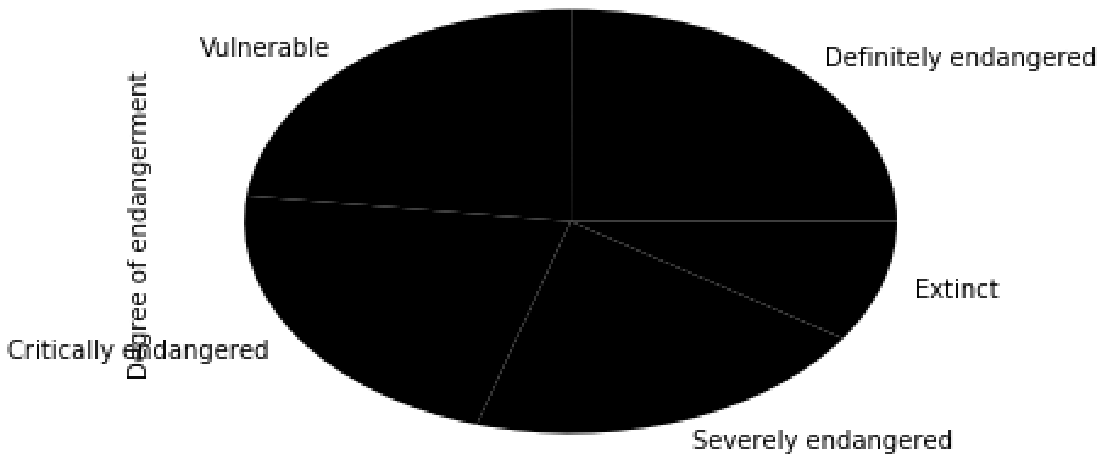


```
In [80]: # [illegible]
         # [illegible]

In [81]: # [illegible]
         # [illegible]
         languages['Degree of endangerment'].v       alue_counts()

Out[81]: Definitely endangered    680
         Vulnerable               628
         Critically endangered    607
         Sev       erely endangered      554
         Extinct                  253
         Name: Degree of endangerment, dtype: int64

In [82]: # [illegible]
         languages['Degree of endangerment'].v       alue_counts().plot(kind='bar'

Out[82]: <matplotlib.axes._subplots.AxesSubplot at 0xc78b48fd30>
```



```
In [84]: # [illegible]
         #[illegible]
         languages['Degree of endangerment'].v       alue_counts().plot(kind='pie'
```

```
Out[84]: <matplotlib.axes._subplots.AxesSubplot at 0xc78a9f7518>
```




```
In [85]: # [illegible]
         critically_endangered=languages[languages['Degree of endangerment'] == 'Cr

In [6]: # [illegible]
        # [illegible]

In [16]: # [illegible]
         ce = languages['Degree of endangerment'] == 'Critically endangered'
         #[illegible]
         v       u = languages['Degree of endangerment'] == 'Vulnerable'
         endangered_languages = languages[ce | v       u]
         #[illegible]

In [91]: # [illegible]
         endangered_languages.to_csv       ('./endangered.csv       ')

In [92]: # [illegible]
```

```
'less' is not recognized as an internal or external command,
operable program or batch file.
```

** Dodatni materijali: **

- https://github.com/jvns/pandas-cookbook
- knjiga : Python for Data Analysis, Wes McKinney

In [ ]: