wyklad-1-1

May 10, 2023

1 Wprowadzenie do przetwarzania danych w języku Python

1.1 Plan wykładu

1. Wprowadzenie do języka Python
2. Wprowadzenie do środowiska Jupyter
3. Wprowadzenie do pakietu Pandas
4. Podstawowe operacje na ramkach danych pakietu Pandas
5. Podsumowanie

2 Python

Python to wysokopoziomowy język programowania z dynamicznym typowaniem. Szczyci się wysoką czytelnością kodu, rozbudowanym ekosystemem bibliotek oraz niskim progiem wejścia. Jest to jeden z najpopularniejszych języków programowania ogólnego zastosowania, popularnym szczególnie w domenach aplikacji webowych, przetwarzania danych oraz uczenia maszynowego.

2.0.1 Przykłady kodu

Deklaracja zmiennych nie wymaga dodatkowych słów kluczowych.

[1]:
```
variable = 42
variable
```

[1]:
```
42
```

Do wbudowanych typów należą m.in. liczby całkowite, liczby zmiennoprzecinkowe, wartości boolowskie, ciągi znaków (`str`, `bytes`), listy, krotki (`tuple`), zbiory oraz słowniki.

[2]:
```
null = None
string = "Example"
integer = 42
boolean = False

print(null, string, integer, boolean)
```

```
None Example 42 False
```

[3]:
```
float_list = [1.0, 2.0, 2.0]
float_list[2] = 3.0
int_tuple = (1, 2)
string_set = {"a", "a", "b"}
dictionary = {"key": "value"}

print(float_list, int_tuple, string_set, dictionary)
```

```
[1.0, 2.0, 3.0] (1, 2) {'b', 'a'} {'key': 'value'}
```

2.0.2 Instrukcje warunkowe

Słowa kluczowe do instrukcji warunkowych to `if`, `elif` oraz `else`.

Wcięcia w Pythonie są częścią składni i wyznaczają zakres instrukcji warunkowych.

[4]:
```
first = None
second = False
if first:
    print("First value was truthy.")
elif second:
    print("Second value was truthy.")
else:
    print("Both values were falsy.")
```

```
Both values were falsy.
```

2.0.3 Pętle

Instrukcje `while` oraz `for` (iteracyjny). `break` i `continue` kontrolują przebieg pętli.

[5]:
```
counter = 3
while counter > 0:
    print("Counter:", counter)
    counter -= 1
```

```
Counter: 3
Counter: 2
Counter: 1
```

[6]:
```
for index in range(3):
    print("Iteration:", index)

print("After iteration", index)
```

```
Iteration: 0
Iteration: 1
Iteration: 2
After iteration 2
```

Za pomocą słowa kluczowego `else` można dodać instrukcje, które są wywoływane tylko jeżeli pętla nie zakończy się poprzez `break`.

[7]:
```
for index in range(1, 9):
    if index % 3 == 0 and index % 2 == 0:
        print("Found value:", index)
        break
    else:
        continue
else:
    print("Value not found!")
```

```
Found value: 6
```

2.0.4 Kolekcje

Python udostępnia przydatne funkcjonalności do przetwarzania kolekcji.

Negatywne indeksy pozwalają odczytywanie wartości z końca list i krotek.

[8]:
```
example_list = [0, 1, 2, 3, 4, 5]
example_list[0], example_list[-1]
```

[8]:
```
(0, 5)
```

Tzw. *slices* pozwalają na wyznaczenie podzbioru kolekcji.

[9]:
```
example_list[0:4]  # Second index is non-inclusive.
```

[9]:
```
[0, 1, 2, 3]
```

Tzw. *list comprehension* ułatwia przekształcanie kolekcji.

[10]:
```
[
    element * 3
    for element in example_list
    if element % 2 == 0
]
```

[10]:
```
[0, 6, 12]
```

Python zawiera system obsługi wyjątków za pomocą słów kluczowych takich jak `raise`, `try`, `except` oraz `finally`.

[11]:
```
try:
    raise Exception("Expected!")
    print("This will not be executed.")
except Exception as exception:
    print("Exception occurred:", exception)
finally:
```

```
    print("This will be printed, no matter what.")
```

```
Exception occurred: Expected!
This will be printed, no matter what.
```

2.0.5 Funkcje

Funkcje definiujemy za pomocą słowa kluczowego `def`.

[12]:
```
def add(a, b):
    return a + b


add(2, 3)
```

[12]:
```
5
```

2.0.6 Klasy

Klasy definiujemy za pomocą słowa kluczowego `class`.

[13]:
```
class Adder:
    def __init__(self, value):
        self.value = value

    def add(self, value):
        return self.value + value


Adder(2).add(3)
```

[13]:
```
5
```

Python wspiera opcjonalne typowanie, jednak nie dokonuje weryfikacji typów podczas działania programu. Tzw. *type hints* używane są np. przez narzędzia do statycznej analizy kodu czy IDE.

[14]:
```
def add(a: int, b: int) -> int:
    return a + b


variable: str = "string"
number: int = 42

# This code will run, even though the variables have different declared types:
variable = number
variable
```

[14]:
```
42
```

3 Jupyter

3.1 Jupyter Notebook

Interaktywny edytor tekstowy działający w przeglądarce, który możliwia na uruchamianie pojedynczych fragmentów kodu i wyświetlanie rezultatów. Zmienne są przechowywane w pamięci, co ułatwia pracę ze zbiorami danych. Środowisko graficzne zapewnia też wsparcie dla wizualizacji, jak również interaktywnych kontrolek umożliwiających wpływ na stan zmiennych z poziomu UI.

3.2 Jupyter Lab

IDE działające w przeglądarce z rozbudowanym zestawem funkcjonalności względem klasycznego środowiska Jupyter Notebook. Przydatny do większych projektów.

3.3 Instalacja

Instrukcja w plikach na Teams.

3.3.1 Docker (rekomendowany)

```
docker run -p 8888:8888 -v "$PWD":/home/jovyan/work jupyter/scipy-notebook
```

3.3.2 pip

```
pip install notebook pandas
jupyter notebook
```

3.3.3 Anaconda

```
conda install -c conda-forge notebook pandas
jupyter notebook
```

4 Pandas

Pandas to oprogramowanie w języku Python do analizy i manipulacji danych. W szczególności, Pandas oferuje struktury danych oraz operacje do przetwarzania tabel oraz serii czasowych. Pandas korzysta z pakietu NumPy do wydajnych operacji numerycznych oraz struktur danych.

4.0.1 Kluczowe struktury danych

- `Series`: struktura danych analogiczna do listy wartości.
- `DataFrame`: ramka danych zawierająca serie tworzące dane tabularyczne.

4.0.2 Import biblioteki

[15]:
```
import pandas as pd
```

[16]:
```
%matplotlib inline
```

4.0.3 Inicjalizacja struktur danych

[17]:
```
species = ["Cat", "Dog", "Rabbit", "Cow"]
```

[18]:
```
pd.Series(species)
```

[18]:
```
0       Cat
1       Dog
2    Rabbit
3       Cow
dtype: object
```

[19]:
```
pd.DataFrame(species)
```

[19]:
```
         0
0     Cat
1     Dog
2  Rabbit
3     Cow
```

Inicjalizacja przez przekazanie obiektu `dict`.

[20]:
```
animals = pd.DataFrame({"animal": species})
animals
```

[20]:
```
   animal
0     Cat
1     Dog
2  Rabbit
3     Cow
```

Odczyt kolumny.

[21]:
```
animals["animal"]
```

[21]:
```
0       Cat
1       Dog
2    Rabbit
3       Cow
Name: animal, dtype: object
```

Dodawanie kolumn.

[22]:
```
animals["age"] = [2, 3, 4, 5]
animals
```

[22]:
```
   animal  age
0     Cat    2
1     Dog    3
```

```
2  Rabbit    4
3     Cow    5
```

Odczyt wiersza.

[23]:
```
animals.loc[0]
```

[23]:
```
animal    Cat
age         2
Name: 0, dtype: object
```

Inicjalizacja przez przekazanie listy słowników.

[24]:
```
pd.DataFrame([
    {"name": "Michal", "age": 27},
    {"name": "Marta", "age": 25}
])
```

[24]:
```
     name  age
0  Michal   27
1   Marta   25
```

Inicjalizacja ramki danych z własnym indeksem.

[25]:
```
persons = pd.DataFrame([
    {"name": "Michal", "age": 27},
    {"name": "Marta", "age": 25}
], index=["A", "B"])
persons
```

[25]:
```
     name  age
A  Michal   27
B   Marta   25
```

[26]:
```
persons.iloc[0], persons.loc["A"]
```

[26]:
```
(name    Michal
 age         27
 Name: A, dtype: object,
 name    Michal
 age         27
 Name: A, dtype: object)
```

4.0.4 Import danych

Wczytanie pliku CSV.

[27]:
```
df = pd.read_csv("Titanic.csv")
df
```

[27]:
```
                                                Name PClass    Age     Sex  \
0                       Allen, Miss Elisabeth Walton    1st  29.00  female
1                        Allison, Miss Helen Loraine    1st   2.00  female
2                Allison, Mr Hudson Joshua Creighton    1st  30.00    male
3     Allison, Mrs Hudson JC (Bessie Waldo Daniels)    1st  25.00  female
4                      Allison, Master Hudson Trevor    1st   0.92    male
…                                                ...    …    …     …
1308                             Zakarian, Mr Artun    3rd  27.00    male
1309                          Zakarian, Mr Maprieder    3rd  26.00    male
1310                               Zenni, Mr Philip    3rd  22.00    male
1311                               Lievens, Mr Rene    3rd  24.00    male
1312                                 Zimmerman, Leo    3rd  29.00    male

      Survived
0            1
1            0
2            0
3            0
4            1
…          …
1308         0
1309         0
1310         0
1311         0
1312         0

[1313 rows x 5 columns]
```

Podgląd danych.

[28]:
```
df.head()
```

[28]:
```
                                    Name PClass    Age     Sex  \
0                  Allen, Miss Elisabeth Walton    1st  29.00  female
1                   Allison, Miss Helen Loraine    1st   2.00  female
2           Allison, Mr Hudson Joshua Creighton    1st  30.00    male
3  Allison, Mrs Hudson JC (Bessie Waldo Daniels)    1st  25.00  female
4                 Allison, Master Hudson Trevor    1st   0.92    male

   Survived
0         1
1         0
2         0
3         0
```

```
4         1
```

[29]:
```
df.tail(10)
```

[29]:
```
                        Name PClass   Age     Sex  Survived
1303      Yasbeck, Mr Antoni    3rd  27.0    male         0
1304     Yasbeck, Mrs Antoni    3rd  15.0  female         1
1305      Youssef, Mr Gerios    3rd   NaN    male         0
1306     Zabour, Miss Hileni    3rd   NaN  female         0
1307     Zabour, Miss Tamini    3rd   NaN  female         0
1308      Zakarian, Mr Artun    3rd  27.0    male         0
1309  Zakarian, Mr Maprieder    3rd  26.0    male         0
1310        Zenni, Mr Philip    3rd  22.0    male         0
1311        Lievens, Mr Rene    3rd  24.0    male         0
1312          Zimmerman, Leo    3rd  29.0    male         0
```

[30]:
```
df.sample()
```

[30]:
```
                                  Name PClass  Age     Sex  Survived
96  Flegenheim, Mrs Alfred (Antoinette)    1st  NaN  female         1
```

[31]:
```
df.sample(20)
```

[31]:
```
                                              Name PClass   Age     Sex  \
138                       Hilliard, Mr Herbert Henry    1st   NaN    male
1161                                Roth, Miss Sarah    3rd   NaN  female
64      Clark, Mrs Walter Miller (Virginia McDowell)    1st  26.0  female
1136                    Pickard (Trembisky), Mr Berk    3rd   NaN    male
941                                Lahowd, Mr Sarkis    3rd   NaN    male
1181                          Salander, Mr Karl Johan    3rd  21.0    male
911                        Karlsson, Mr Eir Gervasius    3rd  21.0    male
1021                                  Mirko, Mr Dika    3rd   NaN    male
210                       Romaine, Mr Charles Hallace    1st   NaN    male
219   Ryerson, Mrs Arthur Larned (Emily Maria Borie)    1st  48.0  female
781                               Drazonovic, Mr Josef    3rd   NaN    male
1186                                Samaan, Mr Youssef    3rd   NaN    male
849                         Hansen, Mr Henry Damsgaard    3rd  21.0    male
466                                  Keane, Mr Daniel    2nd   NaN    male
371                       Coleridge, Mr Regild Charles    2nd  29.0    male
552                    Shelley, Mrs William (Imanita)    2nd  25.0  female
451                               Hosono, Mr Masafumi    2nd  41.0    male
842               Hagland, Mr Konrad Mathias Reiersen    3rd   NaN    male
1283                Vestrom, Miss Hulda Amanda Adolfi    3rd  14.0  female
71                          Crafton, Mr John Bertram    1st   NaN    male

      Survived
138          0
```

```
1161          1
64            1
1136          1
941           0
1181          0
911           1
1021          0
210           1
219           1
781           0
1186          0
849           0
466           0
371           0
552           1
451           1
842           0
1283          0
71            0
```

Informacje o zbiorze danych Wymiary (wiersze, kolumny).

[32]:
```
df.shape
```

[32]:
```
(1313, 5)
```

Dokładne informacje, zużycie pamięci.

[33]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1313 entries, 0 to 1312
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Name      1313 non-null   object
 1   PClass    1312 non-null   object
 2   Age       756 non-null    float64
 3   Sex       1313 non-null   object
 4   Survived  1313 non-null   int64
dtypes: float64(1), int64(1), object(3)
memory usage: 51.4+ KB
```

Statystyki kolumn numerycznych.

[34]:
```
df.describe()
```

[34]:
```
                Age     Survived
count   756.000000  1313.000000
mean     30.397989     0.342727
std      14.259049     0.474802
min       0.170000     0.000000
25%      21.000000     0.000000
50%      28.000000     0.000000
75%      39.000000     1.000000
max      71.000000     1.000000
```

Statystyki poszczególnych kolumn.

[35]:
```
column = df["Age"]
column.count(), column.sum()
```

[35]:
```
(756, 22980.88)
```

[36]:
```
column.mean(), column.median()
```

[36]:
```
(30.397989417989418, 28.0)
```

[37]:
```
column.quantile(0.75)
```

[37]:
```
39.0
```

Unikalne wartości kolumny.

[38]:
```
len(column.unique())
```

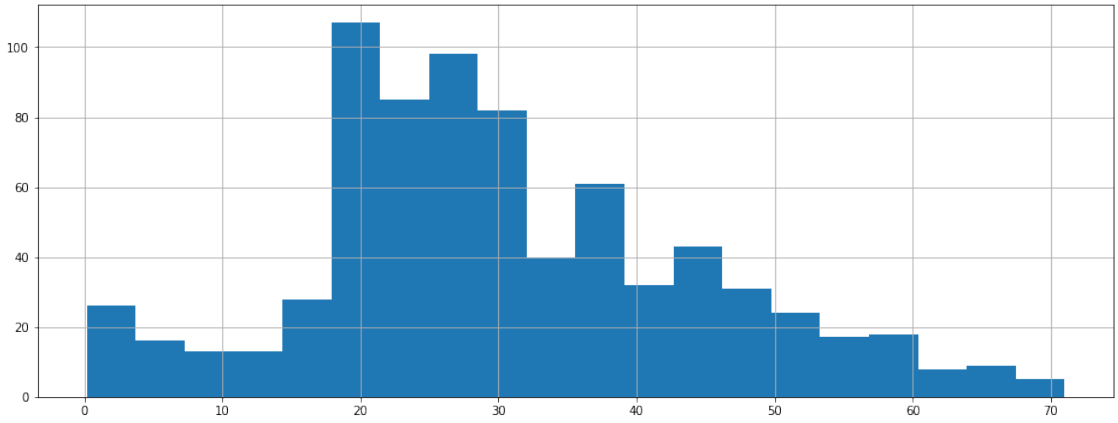
[38]:
```
76
```

[39]:
```
df["PClass"].value_counts()
```

[39]:
```
3rd    711
1st    322
2nd    279
Name: PClass, dtype: int64
```

[40]:
```
df["Age"].hist(bins=20, figsize=(16, 6))
```

[40]:
```
<AxesSubplot:>
```



4.1 Przetwarzanie danych

Obsługa wartości brakujących poprzez usunięcie wierszy z brakującymi danymi.

[41]:
```
df.dropna(subset=["PClass", "Age"])
```

[41]:
```
                                                Name PClass    Age     Sex  \
0                       Allen, Miss Elisabeth Walton    1st  29.00  female
1                        Allison, Miss Helen Loraine    1st   2.00  female
2                Allison, Mr Hudson Joshua Creighton    1st  30.00    male
3     Allison, Mrs Hudson JC (Bessie Waldo Daniels)    1st  25.00  female
4                      Allison, Master Hudson Trevor    1st   0.92    male
…                                                ...    …    …       …
1308                             Zakarian, Mr Artun    3rd  27.00    male
1309                          Zakarian, Mr Maprieder    3rd  26.00    male
1310                               Zenni, Mr Philip    3rd  22.00    male
1311                               Lievens, Mr Rene    3rd  24.00    male
1312                                 Zimmerman, Leo    3rd  29.00    male

      Survived
0            1
1            0
2            0
3            0
4            1
…          …
1308         0
1309         0
1310         0
1311         0
1312         0
```

```
[756 rows x 5 columns]
```

Wypełnienie brakujących wartości stałą.

[42]:
```
df.fillna(0)
```

[42]:
```
                                                Name PClass    Age     Sex  \
0                        Allen, Miss Elisabeth Walton    1st  29.00  female
1                         Allison, Miss Helen Loraine    1st   2.00  female
2                 Allison, Mr Hudson Joshua Creighton    1st  30.00    male
3     Allison, Mrs Hudson JC (Bessie Waldo Daniels)    1st  25.00  female
4                       Allison, Master Hudson Trevor    1st   0.92    male
…                                                 …    …     …     …
1308                               Zakarian, Mr Artun    3rd  27.00    male
1309                            Zakarian, Mr Maprieder    3rd  26.00    male
1310                                 Zenni, Mr Philip    3rd  22.00    male
1311                                 Lievens, Mr Rene    3rd  24.00    male
1312                                   Zimmerman, Leo    3rd  29.00    male

      Survived
0            1
1            0
2            0
3            0
4            1
…          …
1308         0
1309         0
1310         0
1311         0
1312         0

[1313 rows x 5 columns]
```

Wypełnienie brakujących wartości medianą.

[43]:
```
df.fillna(df.median(numeric_only=True))
```

[43]:
```
                                                Name PClass    Age     Sex  \
0                        Allen, Miss Elisabeth Walton    1st  29.00  female
1                         Allison, Miss Helen Loraine    1st   2.00  female
2                 Allison, Mr Hudson Joshua Creighton    1st  30.00    male
3     Allison, Mrs Hudson JC (Bessie Waldo Daniels)    1st  25.00  female
4                       Allison, Master Hudson Trevor    1st   0.92    male
…                                                 …    …     …     …
1308                               Zakarian, Mr Artun    3rd  27.00    male
1309                            Zakarian, Mr Maprieder    3rd  26.00    male
1310                                 Zenni, Mr Philip    3rd  22.00    male
```

```
1311                              Lievens, Mr Rene    3rd  24.00    male
1312                                Zimmerman, Leo    3rd  29.00    male

      Survived
0            1
1            0
2            0
3            0
4            1
…          …
1308         0
1309         0
1310         0
1311         0
1312         0

[1313 rows x 5 columns]
```

4.1.1 Sortowanie

[44]:
```
df.sort_values("Age")
```

[44]:
```
                                        Name PClass   Age     Sex  Survived
763     Dean, Miss Elizabeth Gladys (Millve)    3rd  0.17  female         1
751   Danbom, Master Gilbert Sigvard Emanuel    3rd  0.33    male         0
544           Richards, Master George Sidney    2nd  0.80    male         1
616                      Aks, Master Philip    3rd  0.83    male         1
358            Caldwell, Master Alden Gates    2nd  0.83    male         1
…                                        …    …     …       …          …
1300                  Wiseman, Mr Phillippe    3rd   NaN    male         0
1302                      Yalsevac, Mr Ivan    3rd   NaN    male         1
1305                     Youssef, Mr Gerios    3rd   NaN    male         0
1306                    Zabour, Miss Hileni    3rd   NaN  female         0
1307                    Zabour, Miss Tamini    3rd   NaN  female         0

[1313 rows x 5 columns]
```

[45]:
```
df.dropna().sort_values(["Age", "PClass"], ascending=[False, True])
```

[45]:
```
                                             Name PClass    Age     Sex  \
9                          Artagaveytia, Mr Ramon    1st  71.00    male
119                      Goldschmidt, Mr George B    1st  71.00    male
505                    Mitchell, Mr Henry Michael    2nd  71.00    male
72                   Crosby, Captain Edward Gifford    1st  70.00    male
73   Crosby, Mrs Edward Gifford (Catherine Elizabet…    1st  69.00  female
..                                               …    …     …       …
358                    Caldwell, Master Alden Gates    2nd   0.83    male
```

```
616                                Aks, Master Philip    3rd   0.83    male
544                     Richards, Master George Sidney    2nd   0.80    male
751               Danbom, Master Gilbert Sigvard Emanuel    3rd   0.33    male
763                 Dean, Miss Elizabeth Gladys (Millve)    3rd   0.17  female

     Survived
9           0
119         0
505         0
72          0
73          1
..        …
358         1
616         1
544         1
751         0
763         1

[756 rows x 5 columns]
```

4.1.2 Filtrowanie

[46]:
```
df[df["Age"] == 27]
```

[46]:
```
                                                 Name PClass   Age     Sex  \
57                         Chambers, Mr Norman Campbell    1st  27.0    male
63                              Clark, Mr Walter Miller    1st  27.0    male
78                            Daniel, Mr Robert Williams    1st  27.0    male
80                  Davidson, Mrs Thornton (Orian Hays)    1st  27.0  female
87      Douglas, Mrs Frederick Charles (Suzette Baxter)    1st  27.0  female
214   Rothes, the Countess of (Noel Lucy Martha Dyer…    1st  27.0  female
281                             Widener, Mr Harry Elkins    1st  27.0    male
346                                  Bracken, Mr James H    2nd  27.0    male
372                                   Collander, Mr Erik    2nd  27.0    male
493                                 Mantvila, Rev Joseph    2nd  27.0    male
568                               Troutt, Miss Edwi Celia    2nd  27.0  female
570     Turpin, Mrs William John (Dorothy Anne Woncott)    2nd  27.0  female
652                            Attala (Kalil), Mr Solomon    3rd  27.0    male
667                                     Barry, Miss Julia    3rd  27.0  female
733                                          Cor, Mr Ivan    3rd  27.0    male
752                                    Danoff, Mr Yoto    3rd  27.0    male
855                                    Hedman, Mr Oscar    3rd  27.0    male
869                                  Honkanen, Miss Elu    3rd  27.0  female
875                          Ilmakangas, Miss Ida Livija    3rd  27.0  female
878                                 Jansen, Mr Carl Olof    3rd  27.0    male
904                               Jonsson, Nils Hilding    3rd  27.0    male
1299                                     Wirz, Mr Albert    3rd  27.0    male
```

```
1303                              Yasbeck, Mr Antoni    3rd  27.0    male
1308                              Zakarian, Mr Artun    3rd  27.0    male

      Survived
57           1
63           0
78           1
80           1
87           1
214          1
281          0
346          0
372          0
493          0
568          1
570          0
652          0
667          0
733          0
752          0
855          1
869          1
875          0
878          1
904          0
1299         0
1303         0
1308         0
```

[47]:
```
df[~df["Age"].isna()]
```

[47]:
```
                                              Name PClass    Age     Sex  \
0                        Allen, Miss Elisabeth Walton    1st  29.00  female
1                         Allison, Miss Helen Loraine    1st   2.00  female
2                 Allison, Mr Hudson Joshua Creighton    1st  30.00    male
3     Allison, Mrs Hudson JC (Bessie Waldo Daniels)    1st  25.00  female
4                       Allison, Master Hudson Trevor    1st   0.92    male
…                                                 …    …     …     …
1308                              Zakarian, Mr Artun    3rd  27.00    male
1309                           Zakarian, Mr Maprieder    3rd  26.00    male
1310                                Zenni, Mr Philip    3rd  22.00    male
1311                                Lievens, Mr Rene    3rd  24.00    male
1312                                  Zimmerman, Leo    3rd  29.00    male

      Survived
0            1
1            0
```

```
2            0
3            0
4            1
…          …
1308         0
1309         0
1310         0
1311         0
1312         0

[756 rows x 5 columns]
```

Filtrowanie z negacją (~).

[48]:
```
df[~(df["Age"] > 30)]
```

[48]:
```
                                                Name PClass    Age     Sex  \
0                        Allen, Miss Elisabeth Walton    1st  29.00  female
1                         Allison, Miss Helen Loraine    1st   2.00  female
2                 Allison, Mr Hudson Joshua Creighton    1st  30.00    male
3      Allison, Mrs Hudson JC (Bessie Waldo Daniels)    1st  25.00  female
4                       Allison, Master Hudson Trevor    1st   0.92    male
…                                                 …    …    …     …
1308                              Zakarian, Mr Artun    3rd  27.00    male
1309                           Zakarian, Mr Maprieder    3rd  26.00    male
1310                                 Zenni, Mr Philip    3rd  22.00    male
1311                                 Lievens, Mr Rene    3rd  24.00    male
1312                                   Zimmerman, Leo    3rd  29.00    male

      Survived
0            1
1            0
2            0
3            0
4            1
…          …
1308         0
1309         0
1310         0
1311         0
1312         0

[989 rows x 5 columns]
```

Łączenie filtrów (&, |).

[49]:
```
df[((df["Survived"]) | (df["PClass"] == "1st"))]
```

[49]:
```
                                                Name PClass    Age     Sex  \
0                        Allen, Miss Elisabeth Walton    1st  29.00  female
1                         Allison, Miss Helen Loraine    1st   2.00  female
2                 Allison, Mr Hudson Joshua Creighton    1st  30.00    male
3      Allison, Mrs Hudson JC (Bessie Waldo Daniels)    1st  25.00  female
4                       Allison, Master Hudson Trevor    1st   0.92    male
…                                                 …    …     …     …
1279                              Vartunian, Mr David    3rd  22.00    male
1289                    Wennerstrom, Mr August Edvard    3rd    NaN    male
1293                                Wilkes, Mrs Ellen    3rd  45.00  female
1302                                Yalsevac, Mr Ivan    3rd    NaN    male
1304                               Yasbeck, Mrs Antoni    3rd  15.00  female

      Survived
0            1
1            0
2            0
3            0
4            1
…          …
1279         1
1289         1
1293         1
1302         1
1304         1

[579 rows x 5 columns]
```

Filtrowanie z operacjami na stringach.

[50]:
```
df[df["Name"].str.startswith("Z")]
```

[50]:
```
                        Name PClass   Age     Sex  Survived
1306     Zabour, Miss Hileni    3rd   NaN  female         0
1307     Zabour, Miss Tamini    3rd   NaN  female         0
1308      Zakarian, Mr Artun    3rd  27.0    male         0
1309  Zakarian, Mr Maprieder    3rd  26.0    male         0
1310        Zenni, Mr Philip    3rd  22.0    male         0
1312          Zimmerman, Leo    3rd  29.0    male         0
```

4.1.3 Zapis

[51]:
```
df.to_csv("Output.csv", index=False)
```

5 Podsumowanie

- **Python** to wysokopoziomy język programowania popularny do przetwarzania danych.

- **Jupyter** to interaktywne środowisko umożliwiające tworzenie skryptów m.in. w języku Python.
- **Pandas** to biblioteka do przetwarzania danych.

Zadania na PS:

1. Otwórz notebook z wykładu w środowisku Jupyter.
2. Wczytaj zbiór danych `Titanic.csv`.
3. Wyświetl 10 losowych wierszy z ramki danych.
4. Wyświetl liczbę wierszy w ramce danych.
5. Wyświetl szczegółowe informacje o ramce danych.
6. Zastąp brakujące wartości średnią poszczególnych kolumn.
7. Posortuj zbiór danych po wieku malejąco.
8. Znajdź wszystkich pasażerów, którzy:
    1. Byli mężczyznami.
    2. Przeżyli.
    3. Mieli najwyżej 40 lat.
    4. Przeżyli lub mieli poniżej 10 lat.
    5. Mieli `Miss` w imieniu.
9. Zapisz posortowany zbiór danych bez pustych wartości i indeksu do pliku `Processed.csv`.