zbdihd-ps1

May 10, 2023

[23]:
```
import pandas as pd
```

[24]:
```
%matplotlib inline
```

2. Wczytaj zbiór danych Titanic.csv.

[25]:
```
df = pd.read_csv("Titanic.csv")
df
```

[25]:
```
                                                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]
```

3. Wyświetl 10 losowych wierszy z ramki danych.

[26]:
```
df.sample(10)
```

[26]:

	Name	PClass	Age	Sex	Survived
203	Peuchen, Major Arthur Godfrey	1st	52.0	male	1
696	Cacic, Mr Grego	3rd	18.0	male	0
934	Kink, Miss Maria	3rd	22.0	female	0
941	Lahowd, Mr Sarkis	3rd	NaN	male	0
427	Harper, Rev John	2nd	28.0	male	0
293	Bissetti, Ms Amelia	1st	NaN	female	1
1279	Vartunian, Mr David	3rd	22.0	male	1
911	Karlsson, Mr Eir Gervasius	3rd	21.0	male	1
683	Bourke, Miss Mary	3rd	NaN	female	0
288	Young, Miss Marie Grice	1st	36.0	female	1

4. Wyświetl liczbę wierszy w ramce danych.

[27]:
```
df.shape
```

[27]:
```
(1313, 5)
```

5. Wyświetl szczegółowe informacje o ramce danych.

[28]:
```
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
```

6. Zastąp brakujące wartości średnią poszczególnych kolumn.

[29]:
```
df.fillna(df.mean(numeric_only=True))
```

[29]:

	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]
```

7. Posortuj zbiór danych po wieku malejąco.

[30]:
```
df.sort_values(["Age"], ascending=[False])
```

[30]:
```
                                                 Name PClass   Age     Sex
505                          Mitchell, Mr Henry Michael    2nd  71.0    male  \
119                            Goldschmidt, Mr George B    1st  71.0    male
9                                Artagaveytia, Mr Ramon    1st  71.0    male
72                         Crosby, Captain Edward Gifford    1st  70.0    male
73    Crosby, Mrs Edward Gifford (Catherine Elizabet…    1st  69.0  female
…                                                  …    …   …      …
1300                             Wiseman, Mr Phillippe    3rd   NaN    male
1302                                 Yalsevac, Mr Ivan    3rd   NaN    male
1305                               Youssef, Mr Gerios    3rd   NaN    male
1306                              Zabour, Miss Hileni    3rd   NaN  female
1307                              Zabour, Miss Tamini    3rd   NaN  female

      Survived
505          0
119          0
9            0
72           0
73           1
…          …
1300         0
1302         1
```

```
1305         0
1306         0
1307         0

[1313 rows x 5 columns]
```

8. Wypisz jednolinijkowym poleceniem wiersze wszystkich pasażerów, którzy:

Byli mężczyznami.
Przeżyli.
Mieli najwyżej 40 lat.
Przeżyli lub mieli poniżej 10 lat.
Mieli Miss w imieniu.

[44]:
```
df[(df["Sex"] == "male")]
```

[44]:
```
                                     Name PClass    Age   Sex  Survived
2     Allison, Mr Hudson Joshua Creighton    1st  30.00  male         0
4           Allison, Master Hudson Trevor    1st   0.92  male         1
5                       Anderson, Mr Harry    1st  47.00  male         1
7                  Andrews, Mr Thomas, jr    1st  39.00  male         0
9                  Artagaveytia, Mr Ramon    1st  71.00  male         0
…                                      …    …     …     …        …
1308                   Zakarian, Mr Artun    3rd  27.00  male         0
1309               Zakarian, Mr Maprieder    3rd  26.00  male         0
1310                     Zenni, Mr Philip    3rd  22.00  male         0
1311                     Lievens, Mr Rene    3rd  24.00  male         0
1312                       Zimmerman, Leo    3rd  29.00  male         0

[851 rows x 5 columns]
```

[45]:
```
df[(df["Survived"] == 1)]
```

[45]:
```
                                          Name PClass    Age     Sex  \
0                     Allen, Miss Elisabeth Walton    1st  29.00  female  \
4                    Allison, Master Hudson Trevor    1st   0.92    male
5                              Anderson, Mr Harry    1st  47.00    male
6                  Andrews, Miss Kornelia Theodosia    1st  63.00  female
8     Appleton, Mrs Edward Dale (Charlotte Lamson)    1st  58.00  female
…                                             …    …     …     …
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
4            1
5            1
6            1
8            1
…          …
1279         1
1289         1
1293         1
1302         1
1304         1

[450 rows x 5 columns]
```

[46]:
```
df[(df["Age"] < 41)]
```

[46]:
```
                                                 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

[582 rows x 5 columns]
```

[47]:
```
df[(((df["Survived"] == 1) | (df["Age"] < 10)))]
```

[47]:

	Name	PClass	Age	Sex	Survived
0	Allen, Miss Elisabeth Walton	1st	29.00	female	1
1	Allison, Miss Helen Loraine	1st	2.00	female	0
4	Allison, Master Hudson Trevor	1st	0.92	male	1
5	Anderson, Mr Harry	1st	47.00	male	1
6	Andrews, Miss Kornelia Theodosia	1st	63.00	female	1
…	…	…	…	…	…
1279	Vartunian, Mr David	3rd	22.00	male	1
1289	Wennerstrom, Mr August Edvard	3rd	NaN	male	1
1293	Wilkes, Mrs Ellen	3rd	45.00	female	1
1302	Yalsevac, Mr Ivan	3rd	NaN	male	1
1304	Yasbeck, Mrs Antoni	3rd	15.00	female	1

[466 rows x 5 columns]

[48]:
```
df[~(df["Name"].str.find("Miss") == -1)]
```

[48]:

	Name	PClass	Age	Sex	Survived
0	Allen, Miss Elisabeth Walton	1st	29.0	female	1
1	Allison, Miss Helen Loraine	1st	2.0	female	0
6	Andrews, Miss Kornelia Theodosia	1st	63.0	female	1
27	Bonnell, Miss Caroline	1st	30.0	female	1
28	Bonnell, Miss Elizabeth	1st	58.0	female	1
…	…	…	…	…	…
1269	Van der Planke, Miss Augusta	3rd	18.0	female	0
1276	Van Impe, Miss Catharine	3rd	10.0	female	0
1283	Vestrom, Miss Hulda Amanda Adolfi	3rd	14.0	female	0
1306	Zabour, Miss Hileni	3rd	NaN	female	0
1307	Zabour, Miss Tamini	3rd	NaN	female	0

[236 rows x 5 columns]

9. Zapisz posortowany zbiór danych bez pustych wartości i indeksu do pliku Processed.csv

[49]:
```
df.dropna().sort_values("Age").to_csv("Processed.csv", index=False)
```