

01predavanje

August 1, 2025

1 Uvodno predavanje u podatkovnu znanost

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

```
[2]: putnici = pd.read_csv("../..../datasets/titanic.csv")
```

```
[3]: putnici.head()
```

```
[3]:
```

	Survived	Pclass	Name \
0	0	3	Mr. Owen Harris Braund
1	1	1	Mrs. John Bradley (Florence Briggs Thayer) Cum...
2	1	3	Miss. Laina Heikkinen
3	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle
4	0	3	Mr. William Henry Allen

	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
0	male	22.0	1	0	7.2500
1	female	38.0	1	0	71.2833
2	female	26.0	0	0	7.9250
3	female	35.0	1	0	53.1000
4	male	35.0	0	0	8.0500

```
[4]: putnici.tail()
```

```
[4]:
```

	Survived	Pclass	Name	Sex	Age \
882	0	2	Rev. Jozas Montvila	male	27.0
883	1	1	Miss. Margaret Edith Graham	female	19.0
884	0	3	Miss. Catherine Helen Johnston	female	7.0
885	1	1	Mr. Karl Howell Behr	male	26.0
886	0	3	Mr. Patrick Dooley	male	32.0

	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
882	0	0	13.00
883	0	0	30.00
884	1	2	23.45
885	0	0	30.00
886	0	0	7.75

```
[5]: putnici.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 887 entries, 0 to 886
Data columns (total 8 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Survived                             887 non-null    int64
1   Pclass                               887 non-null    int64
2   Name                                 887 non-null    object
3   Sex                                  887 non-null    object
4   Age                                  887 non-null    float64
5   Siblings/Spouses Aboard              887 non-null    int64
6   Parents/Children Aboard              887 non-null    int64
7   Fare                                  887 non-null    float64
dtypes: float64(2), int64(4), object(2)
memory usage: 55.6+ KB
```

```
[6]: putnici.describe()
```

```
[6]:
```

	Survived	Pclass	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
count	887.000000	887.000000	887.000000	887.000000	887.000000	887.000000
mean	0.385569	2.305524	29.471443	0.525366	0.383315	32.30542
std	0.487004	0.836662	14.121908	1.104669	0.807466	49.78204
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	0.000000	2.000000	20.250000	0.000000	0.000000	7.92500
50%	0.000000	3.000000	28.000000	0.000000	0.000000	14.45420
75%	1.000000	3.000000	38.000000	1.000000	0.000000	31.13750
max	1.000000	3.000000	80.000000	8.000000	6.000000	512.32920

```
[7]: putnici["Sex"].describe()
```

```
[7]: count      887
unique        2
top      male
freq         573
Name: Sex, dtype: object
```

```
[8]: putnici.Sex.value_counts()
```

```
[8]: Sex
     male      573
     female    314
     Name: count, dtype: int64
```

```
[9]: putnici.Pclass.value_counts()
```

```
[9]: Pclass
     3      487
     1      216
     2      184
     Name: count, dtype: int64
```

```
[10]: putnici_group_Pclass = putnici.groupby ("Pclass")
```

```
[11]: putnici_group_Pclass.head()
```

```
[11]:
```

	Survived	Pclass	Name \
0	0	3	Mr. Owen Harris Braund
1	1	1	Mrs. John Bradley (Florence Briggs Thayer) Cum...
2	1	3	Miss. Laina Heikkinen
3	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle
4	0	3	Mr. William Henry Allen
5	0	3	Mr. James Moran
6	0	1	Mr. Timothy J McCarthy
7	0	3	Master. Gosta Leonard Palsson
9	1	2	Mrs. Nicholas (Adele Achem) Nasser
11	1	1	Miss. Elizabeth Bonnell
15	1	2	Mrs. (Mary D Kingcome) Hewlett
17	1	2	Mr. Charles Eugene Williams
20	0	2	Mr. Joseph J Fynney
21	1	2	Mr. Lawrence Beesley
23	1	1	Mr. William Thompson Sloper

	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
0	male	22.0	1	0	7.2500
1	female	38.0	1	0	71.2833
2	female	26.0	0	0	7.9250
3	female	35.0	1	0	53.1000
4	male	35.0	0	0	8.0500
5	male	27.0	0	0	8.4583
6	male	54.0	0	0	51.8625
7	male	2.0	3	1	21.0750
9	female	14.0	1	0	30.0708
11	female	58.0	0	0	26.5500

15	female	55.0	0	0	16.0000
17	male	23.0	0	0	13.0000
20	male	35.0	0	0	26.0000
21	male	34.0	0	0	13.0000
23	male	28.0	0	0	35.5000

```
[12]: putnici_group_Pclass["Age"].max()
```

```
[12]: Pclass
1      80.0
2      70.0
3      74.0
Name: Age, dtype: float64
```

```
[13]: putnici_group_Pclass["Age"].min()
```

```
[13]: Pclass
1      0.92
2      0.67
3      0.42
Name: Age, dtype: float64
```

```
[14]: putnici_group_Pclass["Age"].mean()
```

```
[14]: Pclass
1      38.788981
2      29.868641
3      25.188747
Name: Age, dtype: float64
```

```
[15]: putnici_group_Sex = putnici.groupby("Sex")
```

```
[16]: putnici_group_Sex["Survived"].value_counts()
```

```
[16]: Sex      Survived
female 1          233
        0           81
male   0          464
        1          109
Name: count, dtype: int64
```

```
[17]: putnici_group_PS = putnici.groupby(["Pclass", "Sex"])
```

```
[18]: putnici_group_PS ["Fare"].mean()
```

```
[18]: Pclass Sex
1      female 106.125798
      male    67.226127
```

```

2      female    21.970121
      male      19.741782
3      female    16.118810
      male      12.695466
Name: Fare, dtype: float64

```

```
[19]: putnici_group_PS ["Fare"].agg (["count", "mean", "max", "min"])
```

```

[19]:
      count      mean      max      min
Pclass Sex
1  female     94  106.125798  512.3292  25.9292
   male    122   67.226127  512.3292   0.0000
2  female     76   21.970121   65.0000  10.5000
   male    108   19.741782   73.5000   0.0000
3  female    144   16.118810   69.5500   6.7500
   male    343   12.695466   69.5500   0.0000

```

```
[ ]:
```