

08c_streaming

August 1, 2025

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

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

```
[3]: streaming
```

```
[3]:      Unnamed: 0  KorisnikID  Starost_Korisnika  Broj_Pogledanih_Sati_Mjesecno  \
0              0          1001                52                        13.9
1              1          1002                34                         2.5
2              2          1003                26                        36.4
3              3          1004                21                        44.1
4              4          1005                30                        15.6
..           ...          ...                  ...                        ...
195           195          1196                21                        10.2
196           196          1197                20                        25.6
197           197          1198                20                        21.9
198           198          1199                21                        10.9
199           199          1200                33                        37.6
```

```
      Broj_Uredjaja  Tip_Pretplata  Mjesecna_Pretplata_EUR
0              2      Standard      17.48
1              1  Obiteljska      26.62
2              4      Premium      28.42
3              1      Standard      17.06
4              1      Standard      15.47
..           ...          ...                  ...
195             1      Premium      24.11
196             3      Premium      22.99
197             2      Standard      14.80
198             2      Standard      14.47
199             2      Premium      24.08
```

[200 rows x 7 columns]

```
[4]: streaming.describe()
```

```
[4]:
```

	Unnamed: 0	KorisnikID	Starost_Korisnika	\
count	200.000000	200.000000	200.000000	
mean	99.500000	1100.500000	35.475000	
std	57.879185	57.879185	66.405329	
min	0.000000	1001.000000	-5.000000	
25%	49.750000	1050.750000	23.750000	
50%	99.500000	1100.500000	29.000000	
75%	149.250000	1150.250000	35.000000	
max	199.000000	1200.000000	950.000000	

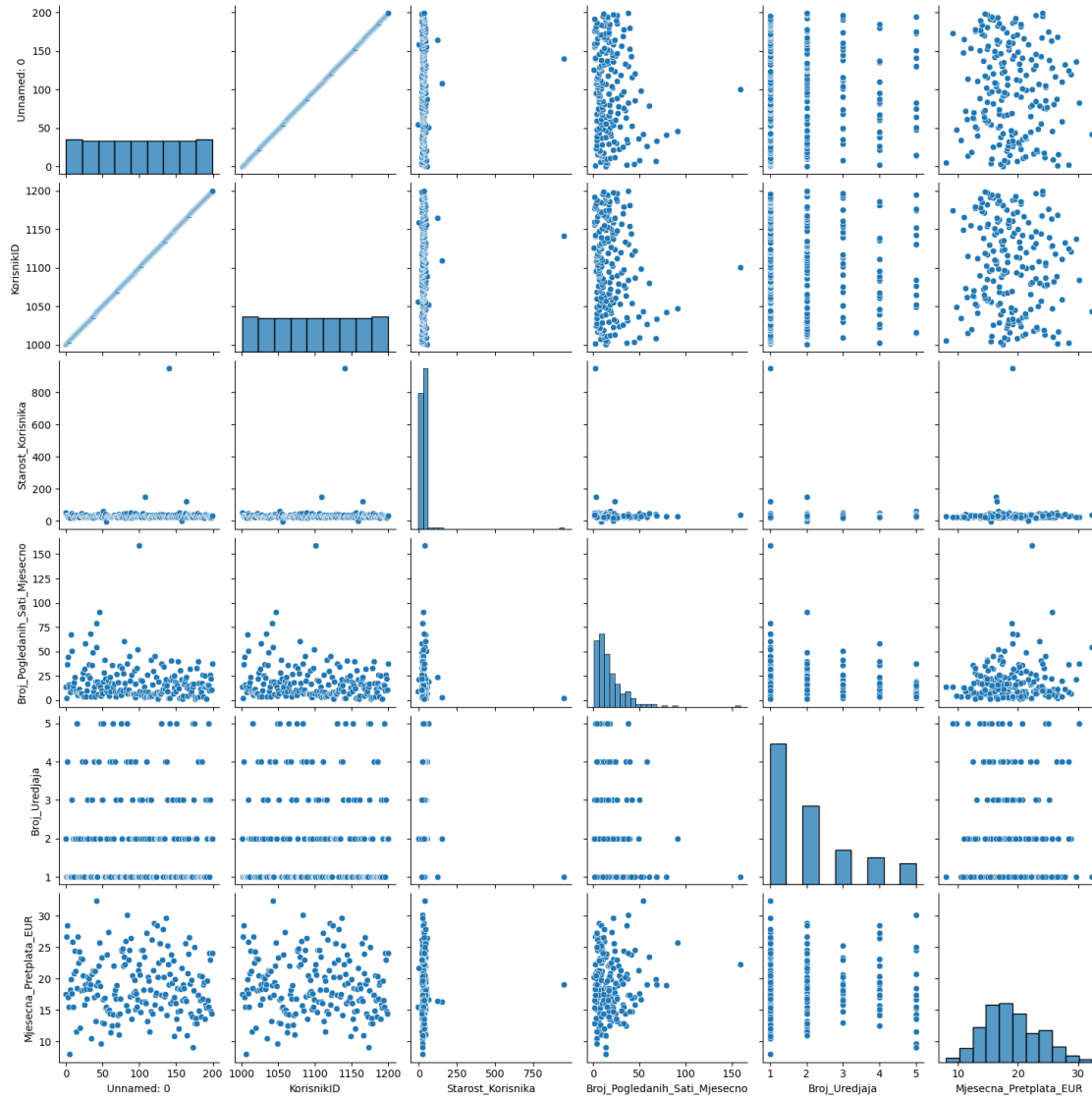
	Broj_Pogledanih_Sati_Mjesecno	Broj_Uredjaja	Mjesecna_Pretplata_EUR
count	200.000000	200.0000	200.000000
mean	18.419500	2.0400	18.833800
std	18.075052	1.2555	4.588201
min	1.000000	1.0000	7.990000
25%	6.875000	1.0000	15.520000
50%	13.750000	2.0000	18.250000
75%	23.525000	3.0000	22.055000
max	158.500000	5.0000	32.340000

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 7 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Unnamed: 0                            200 non-null    int64
1   KorisnikID                            200 non-null    int64
2   Starost_Korisnika                     200 non-null    int64
3   Broj_Pogledanih_Sati_Mjesecno         200 non-null    float64
4   Broj_Uredjaja                         200 non-null    int64
5   Tip_Pretplata                         200 non-null    object
6   Mjesecna_Pretplata_EUR                200 non-null    float64
dtypes: float64(2), int64(4), object(1)
memory usage: 11.1+ KB
```

```
[6]: sns.pairplot (streaming)
```

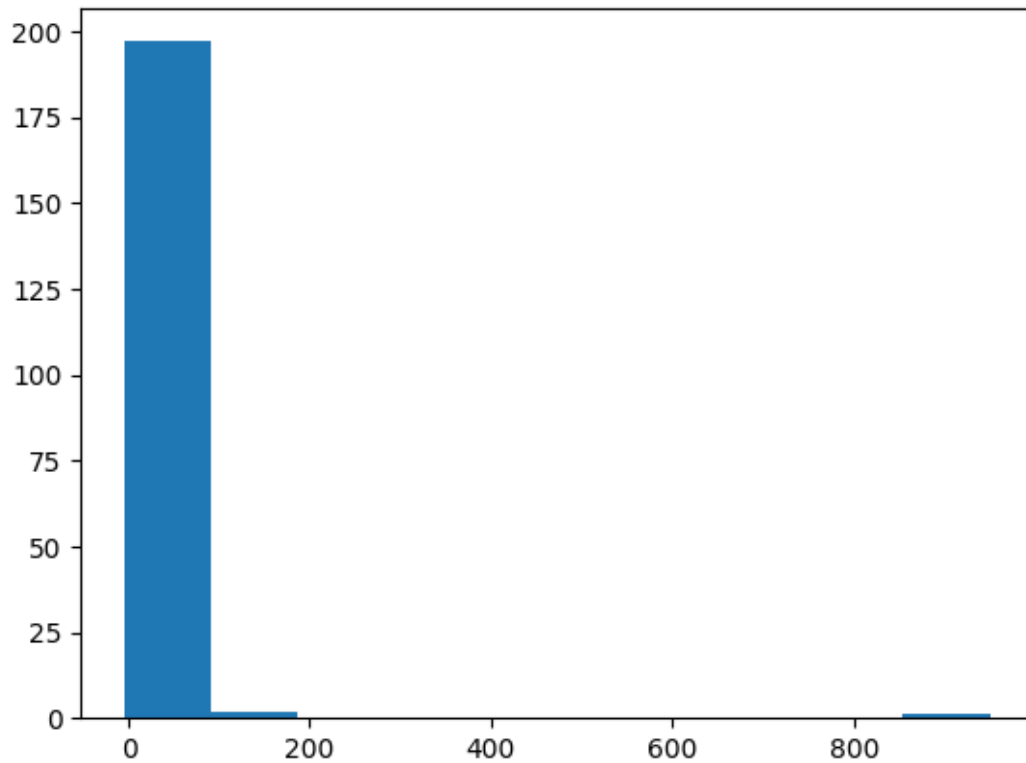
```
[6]: <seaborn.axisgrid.PairGrid at 0x7f7f7157be10>
```



u ovom datasetu očito postoje outliers - riješite ih te istražite kakve su veze medju varijablama (korelacija)

```
[7]: plt.hist(streaming["Starost_Korisnika"])
```

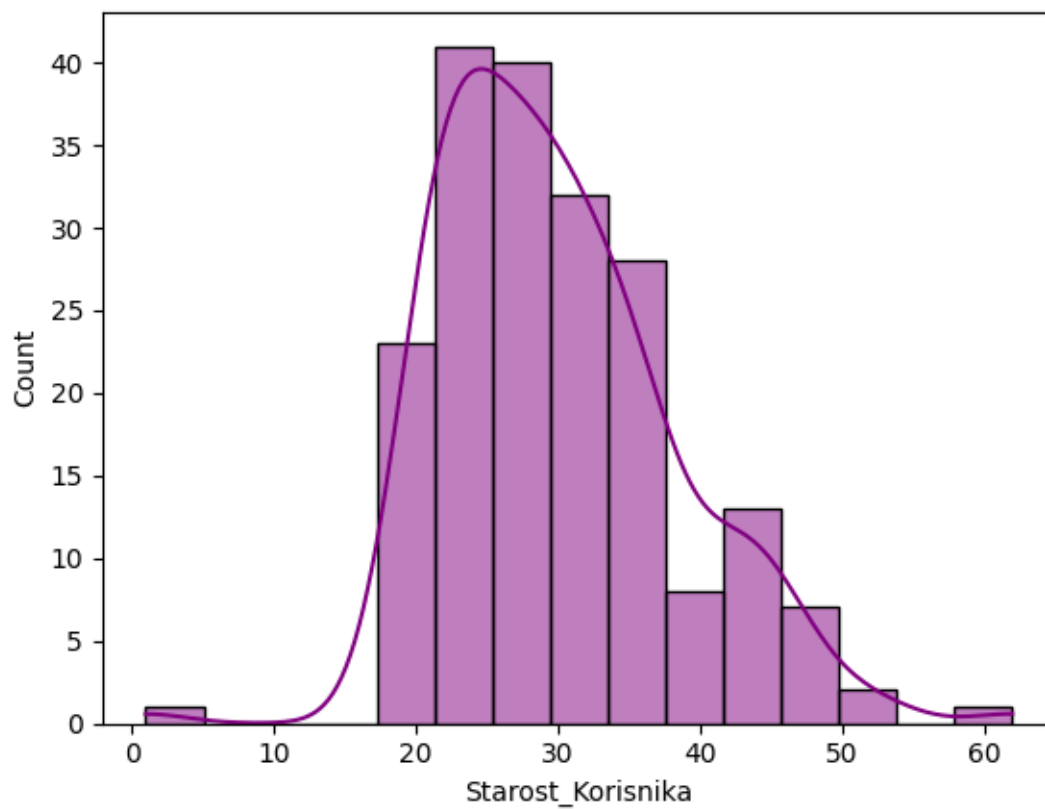
```
[7]: (array([197.,  2.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  1.]),
      array([ -5. ,  90.5, 186. , 281.5, 377. , 472.5, 568. , 663.5, 759. ,
            854.5, 950. ]),
      <BarContainer object of 10 artists>)
```



1 Grafikon u kome je prikazana distribucija starosti korisnika, ali smo makli outliere

```
[8]: sns.histplot(streaming['Starost_Korisnika'][(streaming['Starost_Korisnika'] >= 0) & (streaming['Starost_Korisnika'] <= 100)],
                bins=15, kde=True, color='purple') # kde=True dodaje procjenu
                ↳gustoće distribucije
```

```
[8]: <Axes: xlabel='Starost_Korisnika', ylabel='Count'>
```



[]: