

08a_nekretnine_zagreb

August 1, 2025

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

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

```
[3]: nekretnine_zg.head(10)
```

```
[3]:
```

	Unnamed: 0	Kvart	Tip_Nekretnine	Kvadratura_m2	Broj_Soba	\
0	0	Trešnjevka	Stan	79.0	1	
1	1	Novi Zagreb	Stan	69.0	5	
2	2	Novi Zagreb	Stan	51.0	5	
3	3	Maksimir	Stan	57.0	4	
4	4	Trešnjevka	Poslovni Prostor	74.0	2	
5	5	Centar	Stan	67.0	2	
6	6	Centar	Stan	NaN	15	
7	7	Centar	Poslovni Prostor	89.0	3	
8	8	NaN	Stan	NaN	5	
9	9	Maksimir	Stan	67.0	5	

	Godina_Izgradnje	Cijena_EUR	Datum_Objave_Oglasa
0	1953.028278	70000.0	2023-01-01
1	1967.566714	110000.0	2023-01-02
2	1984.636295	130000.0	2023-01-03
3	1989.460798	187000.0	2023-01-04
4	1984.935049	78000.0	NaN
5	1993.260406	62000.0	2023-01-07
6	NaN	78000.0	NaN
7	1958.518770	100000.0	2023-01-07
8	1964.633673	112000.0	2023-01-07
9	1951.691445	75000.0	2023-01-08

```
[4]: nekretnine_zg.isna().sum()
```

```
[4]:
```

Unnamed: 0	0
Kvart	10
Tip_Nekretnine	0

```

Kvadratura_m2      8
Broj_Soba           0
Godina_Izgradnje    7
Cijena_EUR          0
Datum_Objave_Oglasa 15
dtype: int64

```

```
[5]: nekretnine_zg['Datum_Objave_Oglasa'] = nekretnine_zg['Datum_Objave_Oglasa'].
      ↪ffill()
```

```
[6]: nekretnine_zg.isna().sum()
```

```

[6]: Unnamed: 0      0
Kvart              10
Tip_Nekretnine     0
Kvadratura_m2      8
Broj_Soba           0
Godina_Izgradnje    7
Cijena_EUR          0
Datum_Objave_Oglasa 0
dtype: int64

```

```
[7]: nekretnine_zg.head(10)
```

```

[7]: Unnamed: 0      Kvart  Tip_Nekretnine  Kvadratura_m2  Broj_Soba  \
0         0  Trešnjevka      Stan           79.0          1
1         1   Novi Zagreb      Stan           69.0          5
2         2   Novi Zagreb      Stan           51.0          5
3         3     Maksimir      Stan           57.0          4
4         4  Trešnjevka  Poslovni Prostor       74.0          2
5         5      Centar      Stan           67.0          2
6         6      Centar      Stan           NaN          15
7         7      Centar  Poslovni Prostor       89.0          3
8         8         NaN      Stan           NaN          5
9         9     Maksimir      Stan           67.0          5

```

```

      Godina_Izgradnje  Cijena_EUR  Datum_Objave_Oglasa
0      1953.028278      70000.0      2023-01-01
1      1967.566714     110000.0      2023-01-02
2      1984.636295     130000.0      2023-01-03
3      1989.460798     187000.0      2023-01-04
4      1984.935049      78000.0      2023-01-04
5      1993.260406      62000.0      2023-01-07
6              NaN      78000.0      2023-01-07
7      1958.518770     100000.0      2023-01-07
8      1964.633673     112000.0      2023-01-07
9      1951.691445      75000.0      2023-01-08

```

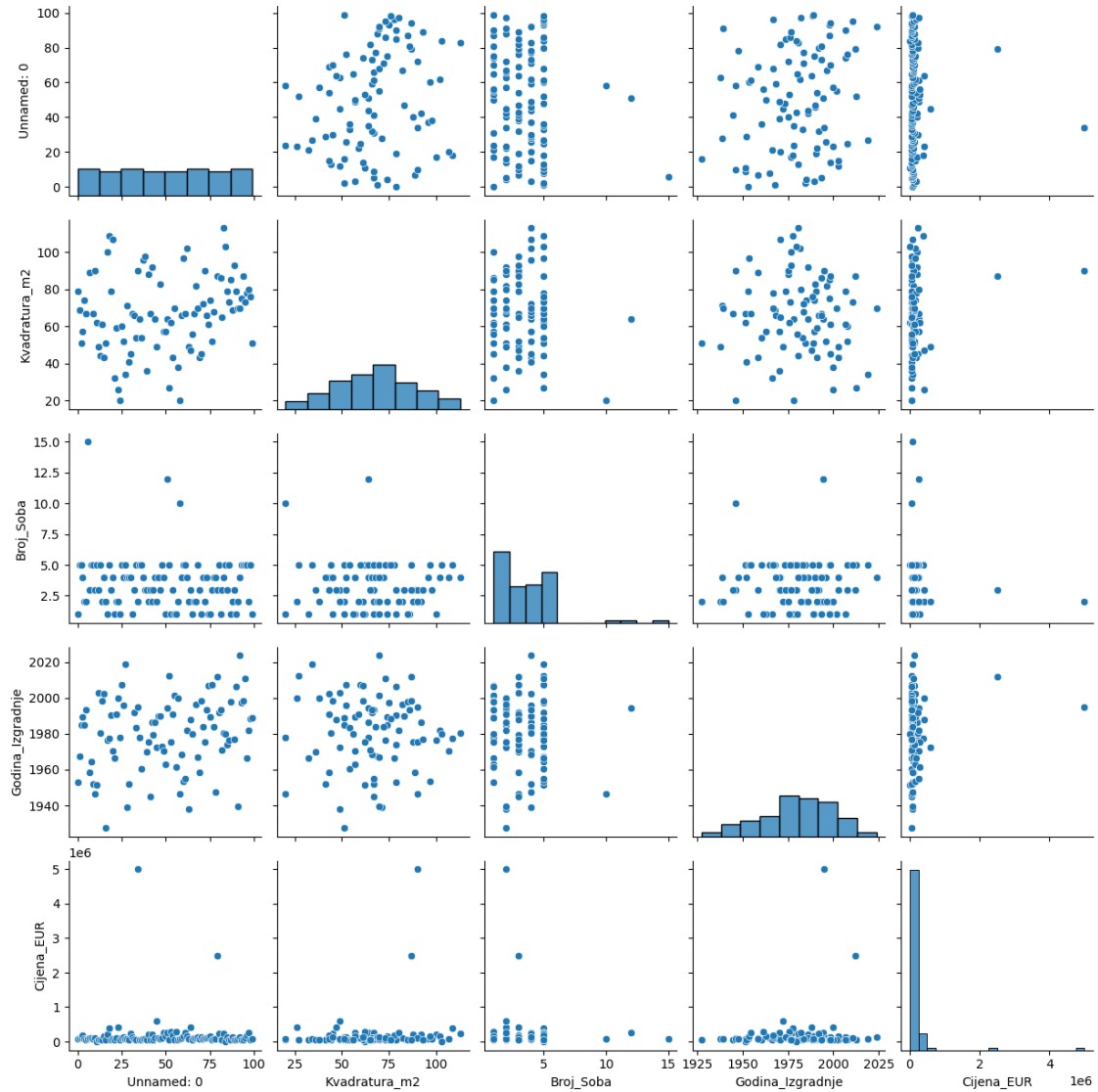
1 čemu služi naredba ffil()?

```
[8]: nekretnine_zg.describe()
```

```
[8]:      Unnamed: 0  Kvadratura_m2  Broj_Soba  Godina_Izgradnje  Cijena_EUR
count  100.000000      92.000000  100.000000      93.000000  1.000000e+02
mean    49.500000      66.956522   3.450000    1979.987316  1.932400e+05
std     29.011492      20.480982   2.152635     19.974325  5.489004e+05
min       0.000000      20.000000   1.000000    1927.476226  1.000000e+03
25%     24.750000      52.000000   2.000000    1967.566714  5.775000e+04
50%     49.500000      67.000000   3.000000    1981.793418  9.050000e+04
75%     74.250000      79.250000   5.000000    1993.508872  1.490000e+05
max     99.000000     113.000000  15.000000    2023.652623  5.000000e+06
```

```
[9]: sns.pairplot(nekretnine_zg)
```

```
[9]: <seaborn.axisgrid.PairGrid at 0x7f4514f24e10>
```



```
[10]: nekretnine_zg.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0            100 non-null   int64
1   Kwart                 90 non-null    object
2   Tip_Nekretnine       100 non-null   object
3   Kvadratura_m2        92 non-null    float64
4   Broj_Soba            100 non-null   int64
5   Godina_Izgradnje     93 non-null    float64
```

```
6  Cijena_EUR          100 non-null    float64
7  Datum_Objave_Oglasa 100 non-null    object
dtypes: float64(3), int64(2), object(3)
memory usage: 6.4+ KB
```

2 zadatak - u ovom datasetu očito ima ekstremnih vrijednosti te outliera... popraviti!

[]: