

02osiguranici

August 1, 2025

1 Analiza osiguranika

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

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

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

```
[3]:
```

	City	Gender	Age	Income	Illness
0	Dallas	Male	41	40367.0	No
1	Dallas	Male	54	45084.0	No
2	Dallas	Male	42	52483.0	No
3	Dallas	Male	40	40941.0	No
4	Dallas	Male	46	50289.0	No

```
[4]: osiguranici.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150000 entries, 0 to 149999
Data columns (total 5 columns):
#   Column      Non-Null Count  Dtype
---  -
0   City         150000 non-null object
1   Gender       150000 non-null object
2   Age          150000 non-null int64
3   Income       150000 non-null float64
4   Illness      150000 non-null object
dtypes: float64(1), int64(1), object(3)
memory usage: 5.7+ MB
```

```
[5]: osiguranici.describe()
```

```
[5]:
```

	Age	Income
count	150000.000000	150000.000000
mean	44.950200	91252.798273
std	11.572486	24989.500948
min	25.000000	-654.000000
25%	35.000000	80867.750000

50%	45.000000	93655.000000
75%	55.000000	104519.000000
max	65.000000	177157.000000

```
[6]: osiguranici.Illness.describe()
```

```
[6]: count      150000
      unique        2
      top         No
      freq      137861
      Name: Illness, dtype: object
```

```
[7]: osiguranici.Illness.value_counts()
```

```
[7]: Illness
     No      137861
     Yes      12139
     Name: count, dtype: int64
```

```
[8]: osiguranici_po_bolesti = osiguranici.groupby(["Illness"])
      osiguranici_po_GB = osiguranici.groupby(["Gender", "Illness"])
```

```
[9]: osiguranici_po_GB.Illness.value_counts()
```

```
[9]: Gender  Illness
     Female No      60869
           Yes      5331
     Male   No      76992
           Yes      6808
     Name: count, dtype: int64
```

```
[10]: osiguranici.Age.value_counts()
```

```
[10]: Age
     26      3918
     64      3879
     43      3862
     41      3862
     54      3838
     29      3805
     28      3798
     27      3790
     63      3784
     44      3782
     36      3780
     59      3775
     47      3773
     37      3771
```

```

33    3769
58    3762
49    3762
42    3760
32    3759
52    3753
31    3749
48    3743
40    3740
61    3737
38    3734
57    3732
30    3731
51    3729
53    3724
34    3710
46    3707
50    3692
55    3687
62    3684
39    3675
45    3670
35    3658
56    3602
60    3582
25    1868
65    1864
Name: count, dtype: int64

```

```
[11]: osiguranici_po_bolesti["Illness"].agg(["count"])
```

```

[11]:          count
Illness
No      137861
Yes      12139

```

```
[12]: osiguranici_po_GB["Gender"].agg(["count"])
```

```

[12]:          count
Gender Illness
Female No      60869
        Yes      5331
Male   No      76992
        Yes      6808

```

```
[13]: osiguranici_po_gradu = osiguranici.groupby("City")
```

```
[14]: osiguranici_po_gradu["Income"].agg(["count", "min", "max", "mean"])
```

```
[14]:
```

	count	min	max	mean
City				
Austin	12292	50602.0	132302.0	90277.513423
Boston	8301	48425.0	132348.0	91554.571497
Dallas	19707	-654.0	91479.0	45252.231187
Los Angeles	32173	51131.0	141249.0	95264.155410
Mountain View	14219	95202.0	177157.0	135078.415782
New York City	50307	55538.0	147170.0	96857.131393
San Diego	4881	64941.0	140190.0	100756.209178
Washington D.C.	8120	33335.0	107704.0	70991.612808

```
[15]: osiguranici_po_gradu["Illness"].describe()
```

```
[15]:
```

	count	unique	top	freq
City				
Austin	12292	2	No	11281
Boston	8301	2	No	7615
Dallas	19707	2	No	18094
Los Angeles	32173	2	No	29605
Mountain View	14219	2	No	13041
New York City	50307	2	No	46286
San Diego	4881	2	No	4487
Washington D.C.	8120	2	No	7452

2 Zaključak

Muškarci više ugovaraju osiguranje, ali su i većinom bolesni. Ukupno je 137861 (60869 Ž, 76992 M) zdrave osobe i 12139 (5331 Ž, 6808 M) bolesnih. Najviše osiguranika živi u NYC, a isto tako ima i najviše bolesnih ljudi.