

SAD_Zdravstvo

August 1, 2025

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

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

```
[3]: health.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  -
 0   age         1338 non-null   int64
 1   sex         1338 non-null   object
 2   bmi         1338 non-null   float64
 3   children    1338 non-null   int64
 4   smoker      1338 non-null   object
 5   region      1338 non-null   object
 6   charges     1338 non-null   float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.3+ KB
```

```
[4]: health.head()
```

```
[4]:   age    sex    bmi  children  smoker    region    charges
0   19  female  27.900         0    yes  southwest  16884.92400
1   18   male  33.770         1    no   southeast   1725.55230
2   28   male  33.000         3    no   southeast   4449.46200
3   33   male  22.705         0    no  northwest  21984.47061
4   32   male  28.880         0    no  northwest   3866.85520
```

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

```
[5]:      count      age      bmi      children      charges
count  1338.000000  1338.000000  1338.000000  1338.000000  1338.000000
mean     39.207025   30.663397    1.094918  13270.422265
std     14.049960    6.098187    1.205493  12110.011237
```

min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.296250	0.000000	4740.287150
50%	39.000000	30.400000	1.000000	9382.033000
75%	51.000000	34.693750	2.000000	16639.912515
max	64.000000	53.130000	5.000000	63770.428010

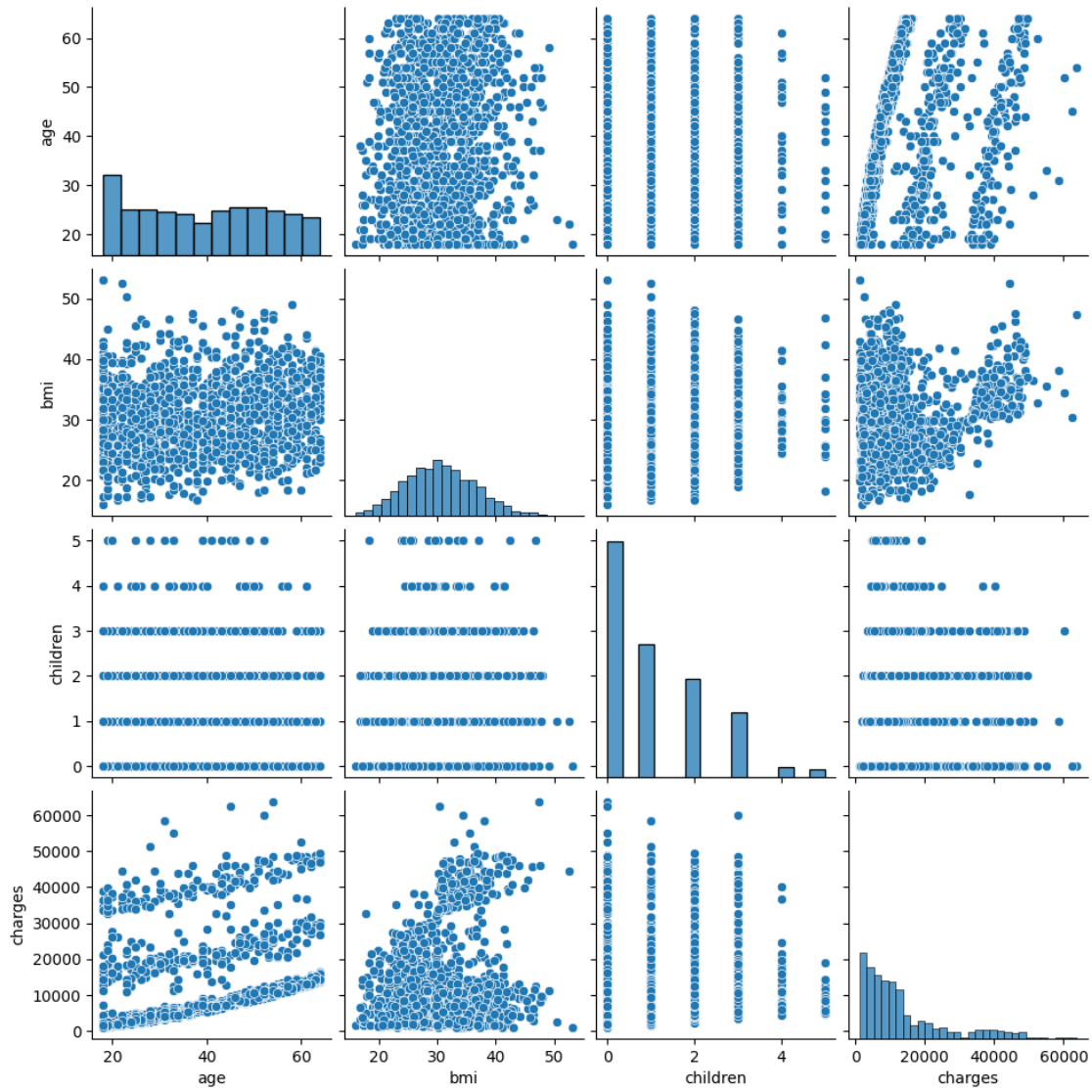
```
[6]: health.corr(numeric_only=True)
```

```
[6]:
```

	age	bmi	children	charges
age	1.000000	0.109272	0.042469	0.299008
bmi	0.109272	1.000000	0.012759	0.198341
children	0.042469	0.012759	1.000000	0.067998
charges	0.299008	0.198341	0.067998	1.000000

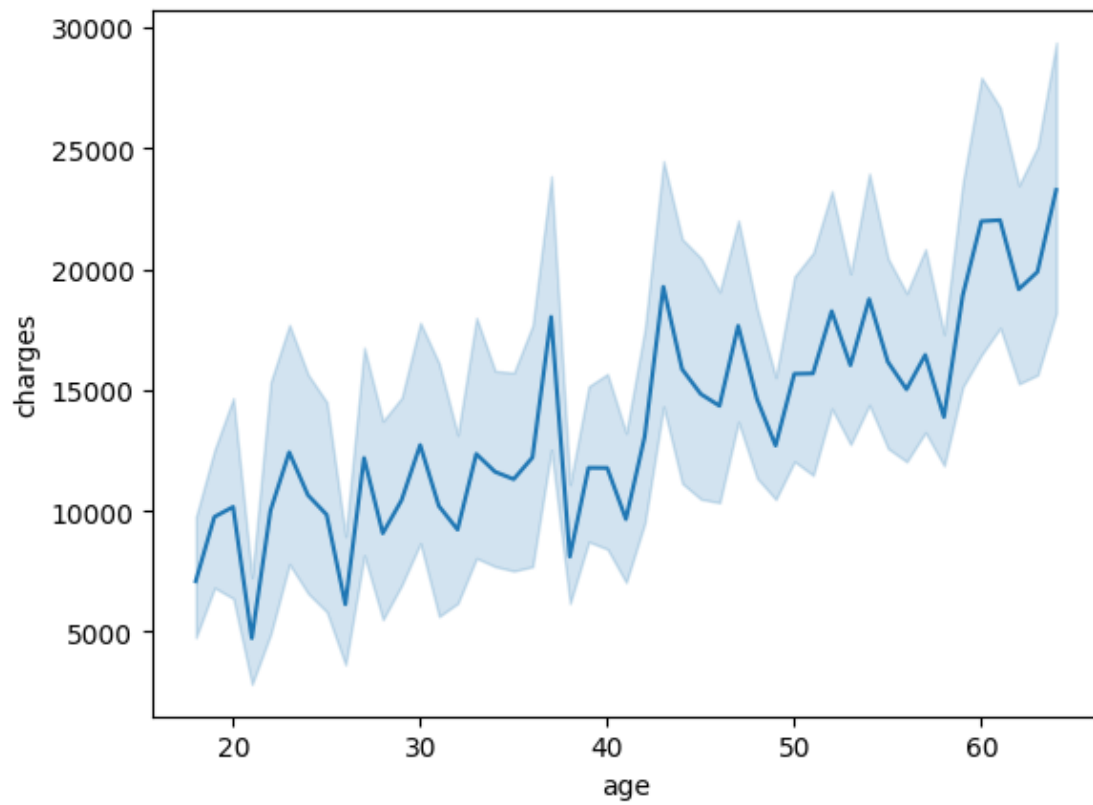
```
[7]: sns.pairplot(health)
```

```
[7]: <seaborn.axisgrid.PairGrid at 0x7f92bf634150>
```



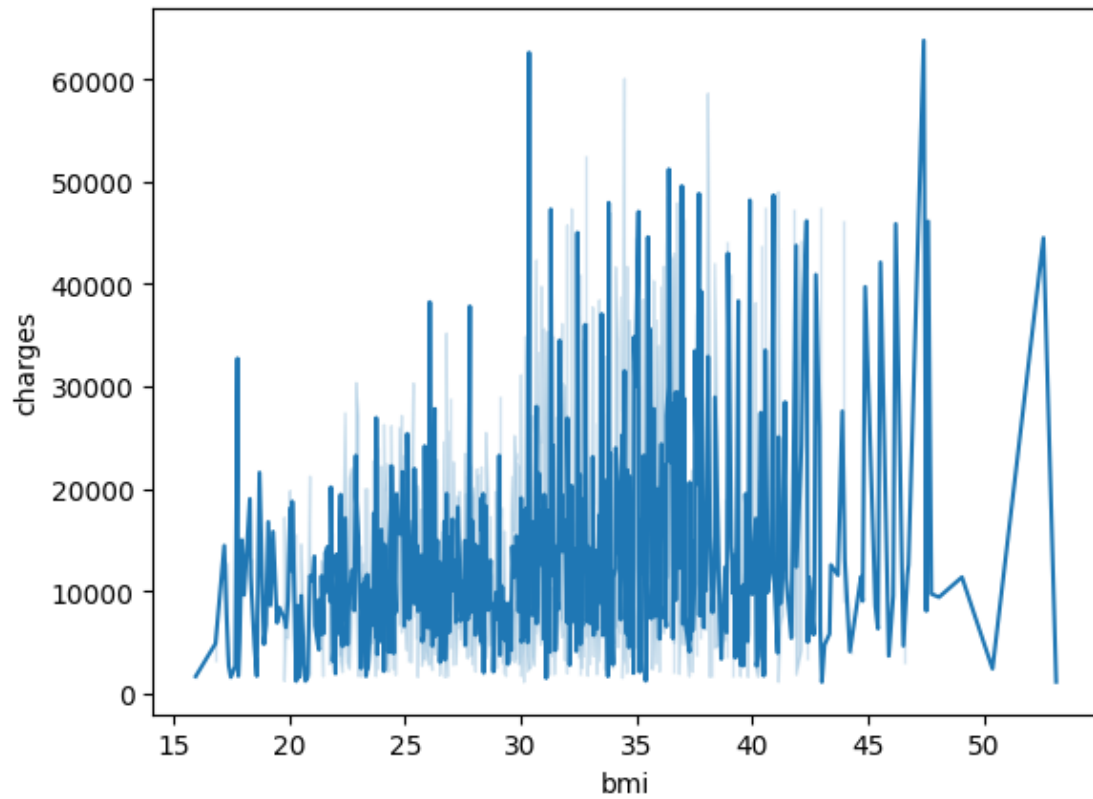
```
[8]: sns.lineplot(health, x=health.age, y=health.charges)
```

```
[8]: <Axes: xlabel='age', ylabel='charges'>
```



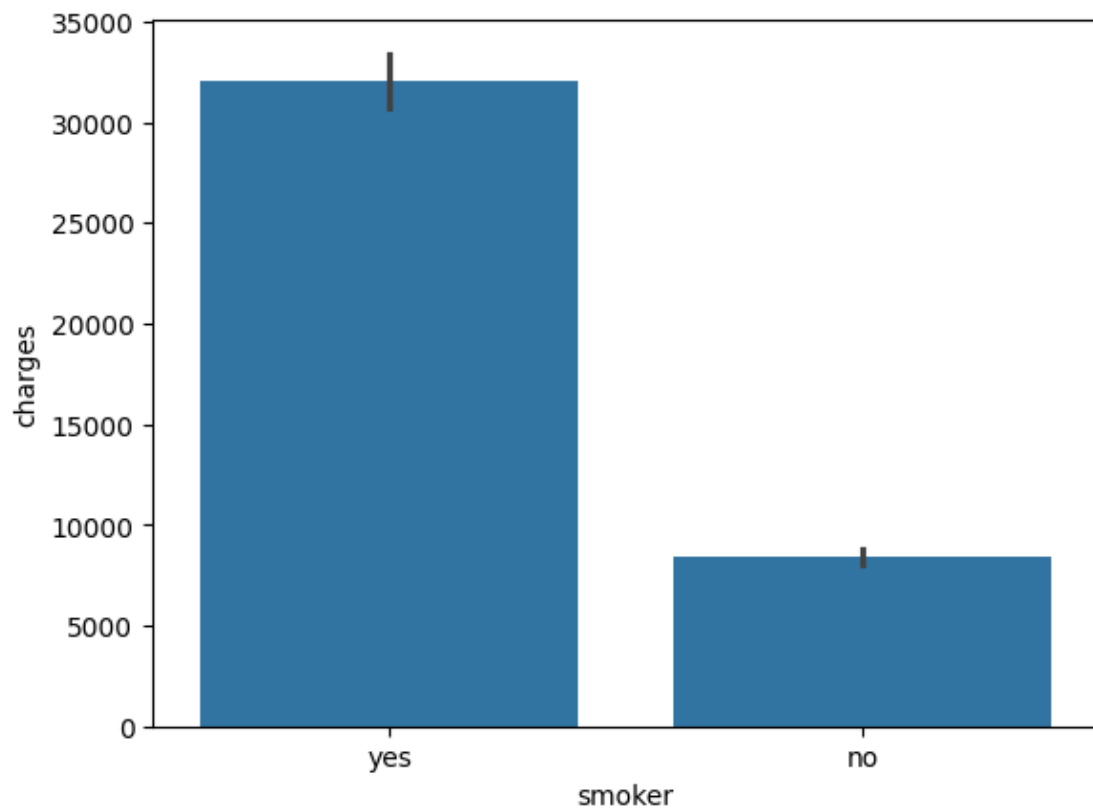
```
[9]: sns.lineplot(health, x=health.bmi, y=health.charges)
```

```
[9]: <Axes: xlabel='bmi', ylabel='charges'>
```



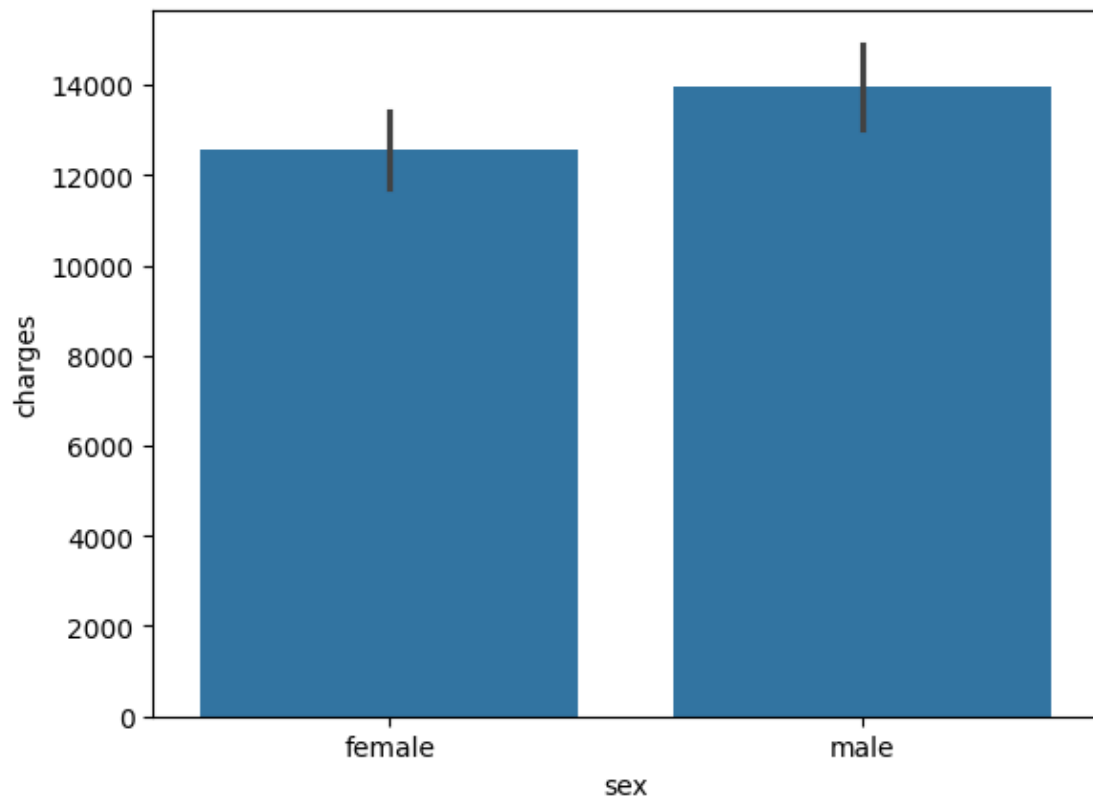
```
[10]: sns.barplot(health, x=health.smoker, y=health.charges)
```

```
[10]: <Axes: xlabel='smoker', ylabel='charges'>
```



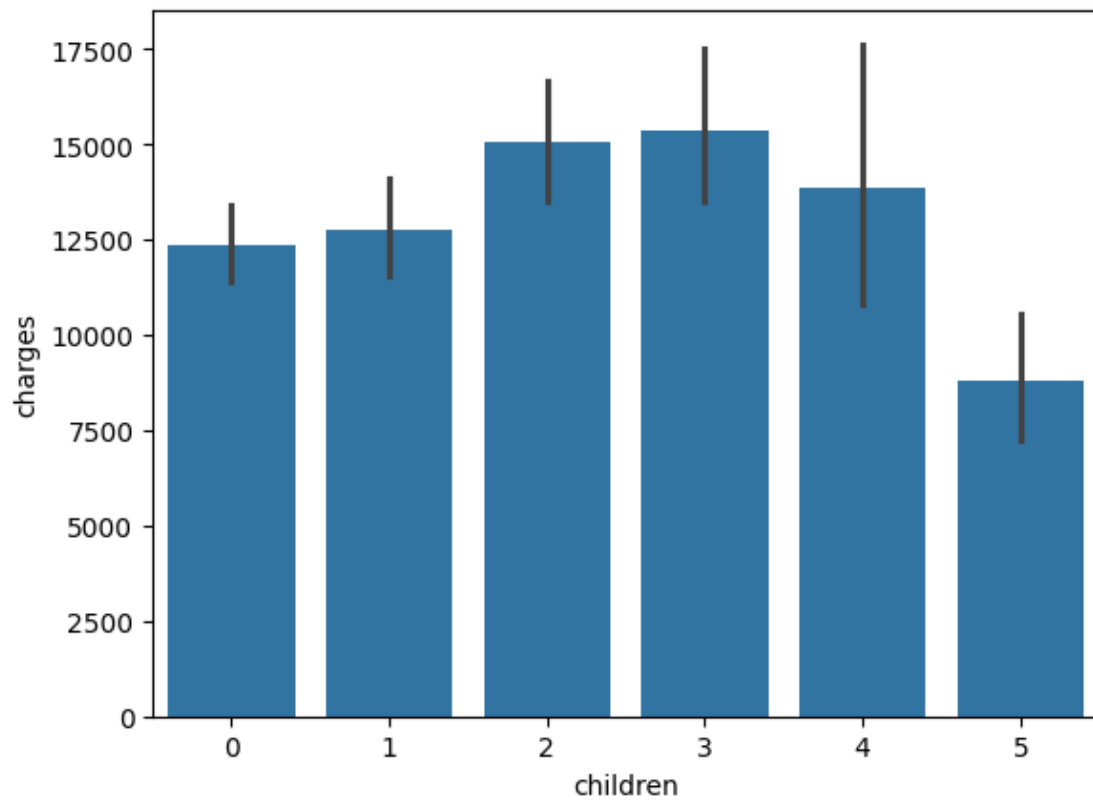
```
[11]: sns.barplot(health, x=health.sex, y=health.charges)
```

```
[11]: <Axes: xlabel='sex', ylabel='charges'>
```



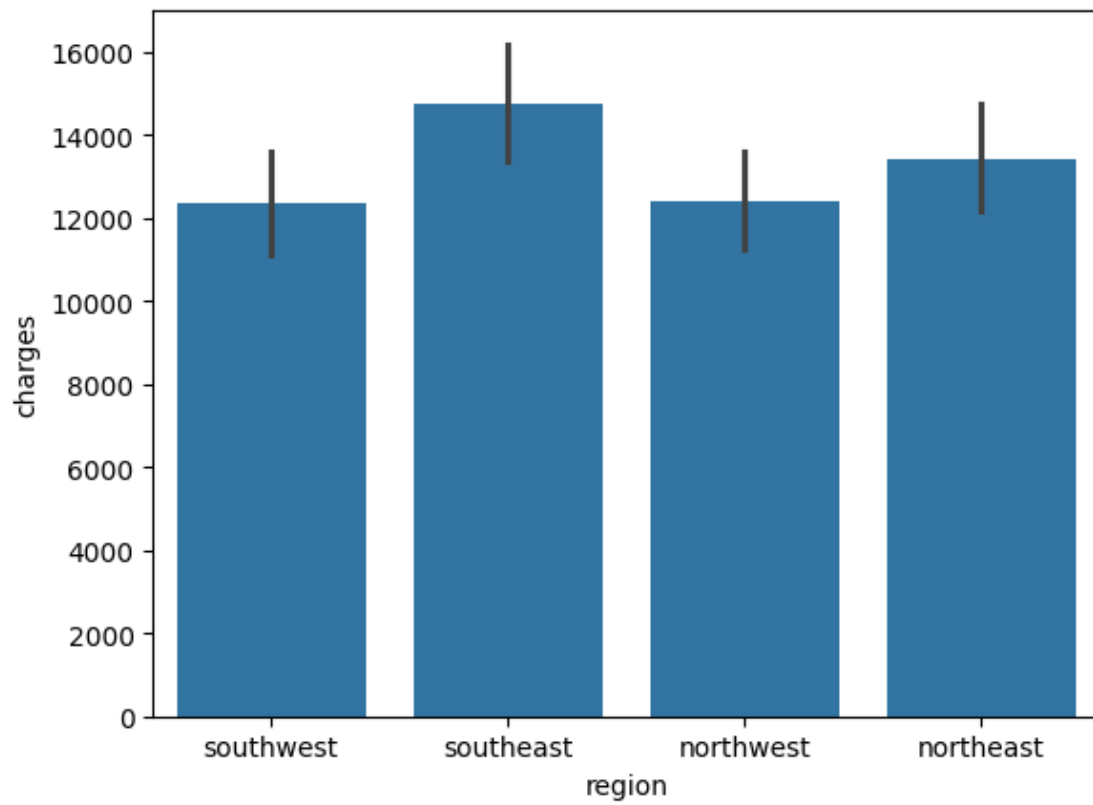
```
[12]: sns.barplot(health, x=health.children, y=health.charges)
```

```
[12]: <Axes: xlabel='children', ylabel='charges'>
```



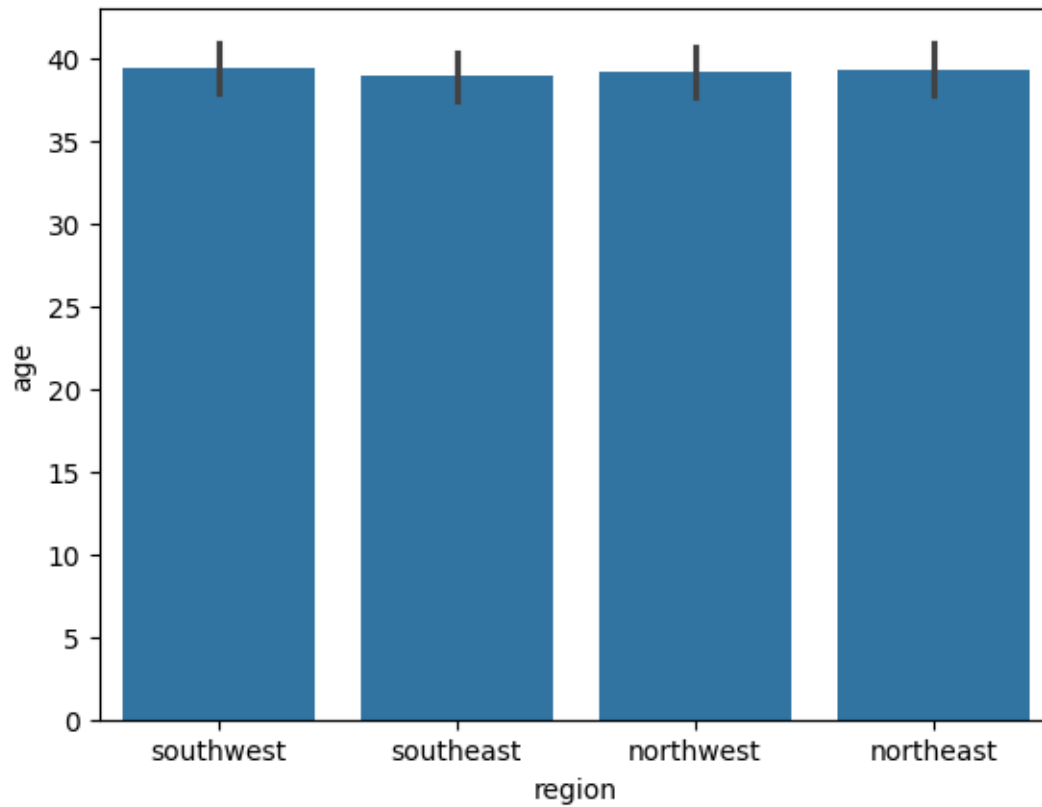
```
[13]: sns.barplot(health, x=health.region, y=health.charges)
```

```
[13]: <Axes: xlabel='region', ylabel='charges'>
```



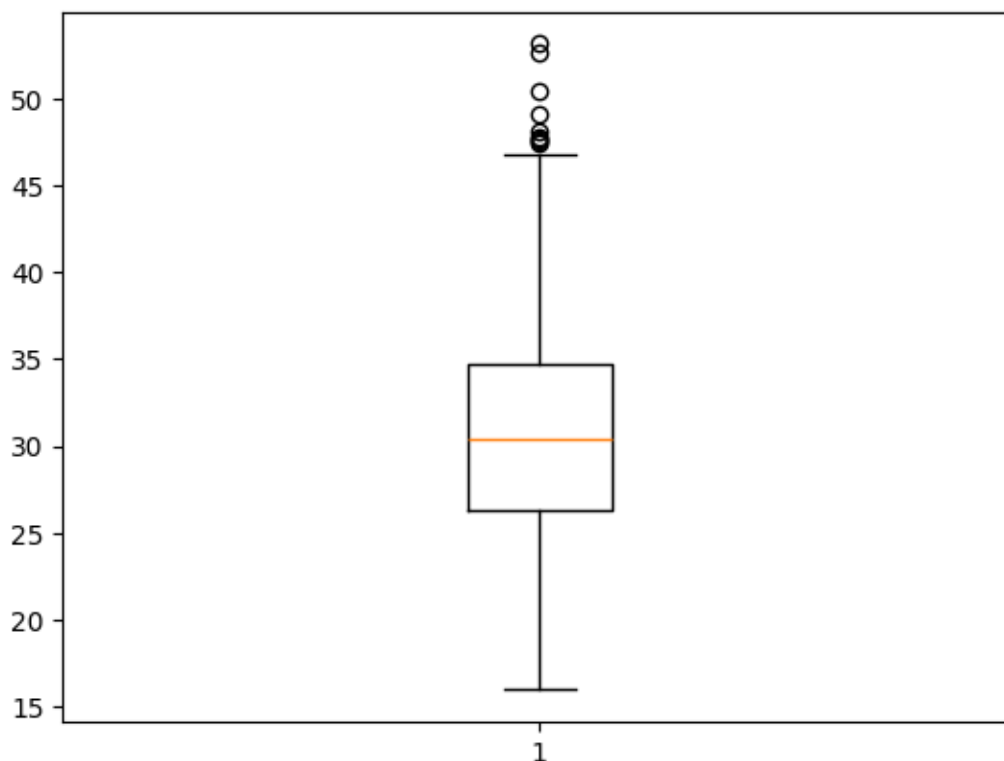
```
[14]: sns.barplot(health, x=health.region, y=health.age)
```

```
[14]: <Axes: xlabel='region', ylabel='age'>
```



```
[15]: plt.boxplot(health["bmi"])
```

```
[15]: {'whiskers': [<matplotlib.lines.Line2D at 0x7f92b9201a10>,
<matplotlib.lines.Line2D at 0x7f92b92025d0>],
'caps': [<matplotlib.lines.Line2D at 0x7f92b91f2010>,
<matplotlib.lines.Line2D at 0x7f92b9203a50>],
'boxes': [<matplotlib.lines.Line2D at 0x7f92b9200e90>],
'medians': [<matplotlib.lines.Line2D at 0x7f92b920c5d0>],
'fliers': [<matplotlib.lines.Line2D at 0x7f92b920d090>],
'means': []}
```



Podaci pokazuju da amerikanci u prosjeku imaju prekomjernu tjelesnu težinu te da postoje anomalije za vrijednosti iznad 45 BMI.

1 Zaključak

- Na cijenu liječenja utječu najviše dob, pušenje i indeks tjelesne mase (BMI).
- Razlika u cijeni liječenja muškaraca i žena je jako mala te spol nije važna varijabla u određivanju konačnih troškova liječenja. Također postoji neznatna razlika u cijeni troškova na istoku i zapadu, gdje je na zapadu cijena liječenja nešto niža. Dob populacije ljudi po regijama je izjednačena.

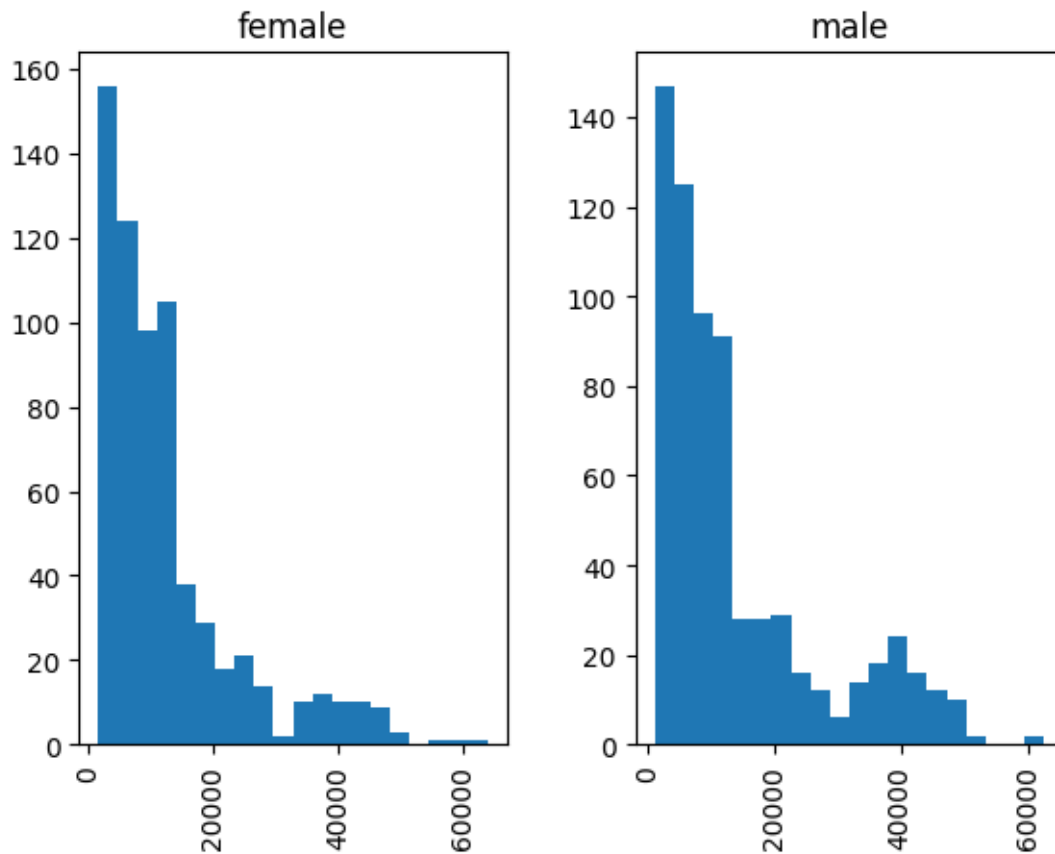
2 Prvi dio zadatka

- Osnovna analiza
- Raspon određenih vrijednosti (BMI)
- Varijable koje utječu na cijenu liječenja
- Korelacije
- Zaključak u obliku izvještaja - Varijable koje utječu na cijenu liječenja
- `value_counts()` za varijable
- grupiranje

3 Histograms

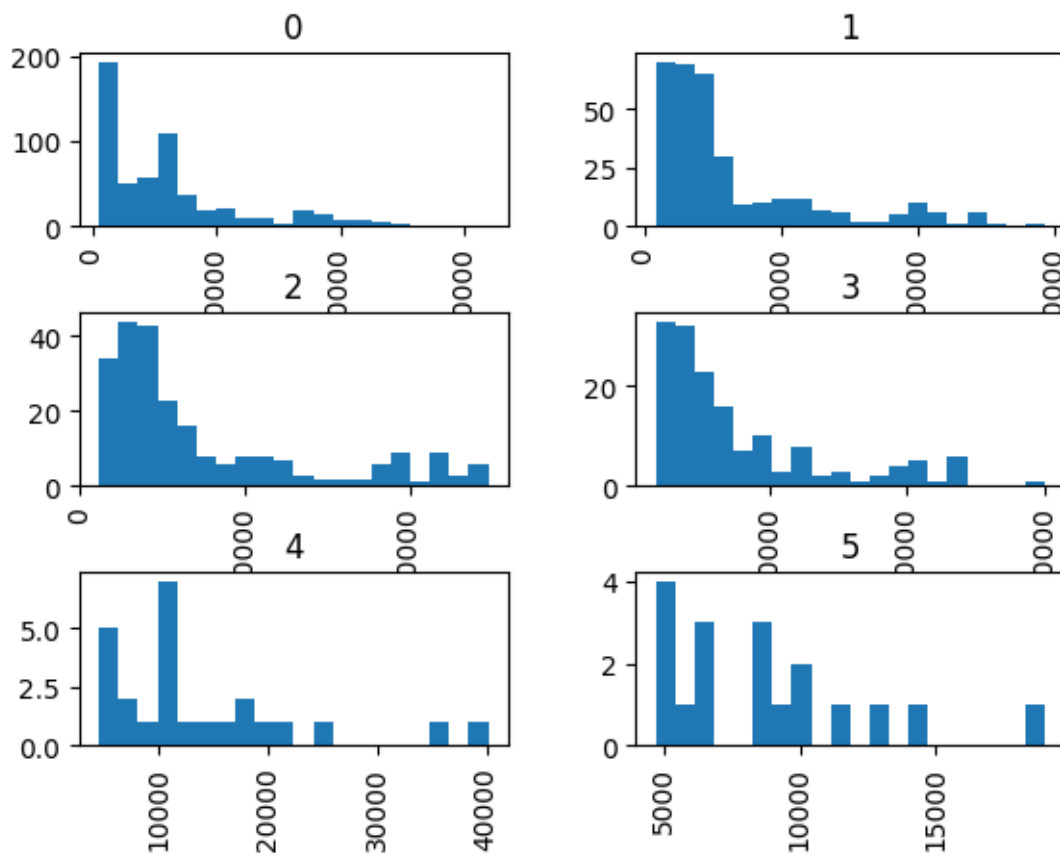
```
[16]: health["charges"].hist(by = health["sex"], bins = 20)
```

```
[16]: array([<Axes: title={'center': 'female'}>,
        <Axes: title={'center': 'male'}>], dtype=object)
```



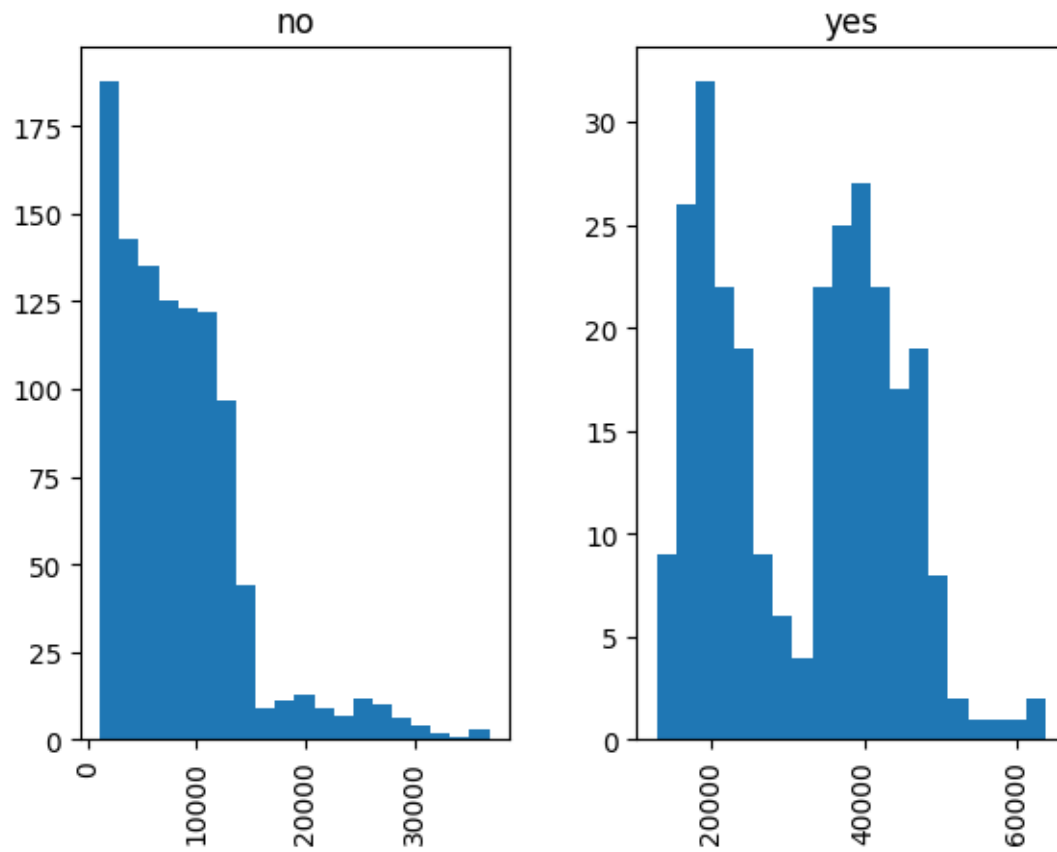
```
[17]: health["charges"].hist(by = health["children"], bins = 20)
```

```
[17]: array([[<Axes: title={'center': '0'}>, <Axes: title={'center': '1'}>],
        [<Axes: title={'center': '2'}>, <Axes: title={'center': '3'}>],
        [<Axes: title={'center': '4'}>, <Axes: title={'center': '5'}>]],
        dtype=object)
```



```
[18]: health["charges"].hist(by = health["smoker"], bins = 20)
```

```
[18]: array([<Axes: title={'center': 'no'}>, <Axes: title={'center': 'yes'}>],
      dtype=object)
```



```
[19]: health["charges"].hist(by = health["region"], bins = 20)
```

```
[19]: array([[<Axes: title={'center': 'northeast'}>,
          <Axes: title={'center': 'northwest'}>],
          [<Axes: title={'center': 'southeast'}>,
          <Axes: title={'center': 'southwest'}>]], dtype=object)
```

