Brain_stroke

Импортируем необходимые библиотеки и выведем данные.

In [12]:

```
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv('brain_stroke.csv')
df.dropna(inplace=True)

df.head(10)
```

Out[12]:

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_
0	Male	67.0	0	1	Yes	Private	Urban	
1	Male	80.0	0	1	Yes	Private	Rural	
2	Female	49.0	0	0	Yes	Private	Urban	
3	Female	79.0	1	0	Yes	Self-employed	Rural	
4	Male	81.0	0	0	Yes	Private	Urban	
5	Male	74.0	1	1	Yes	Private	Rural	
6	Female	69.0	0	0	No	Private	Urban	
7	Female	78.0	0	0	Yes	Private	Urban	
8	Female	81.0	1	0	Yes	Private	Rural	
9	Female	61.0	0	1	Yes	Govt_job	Rural	

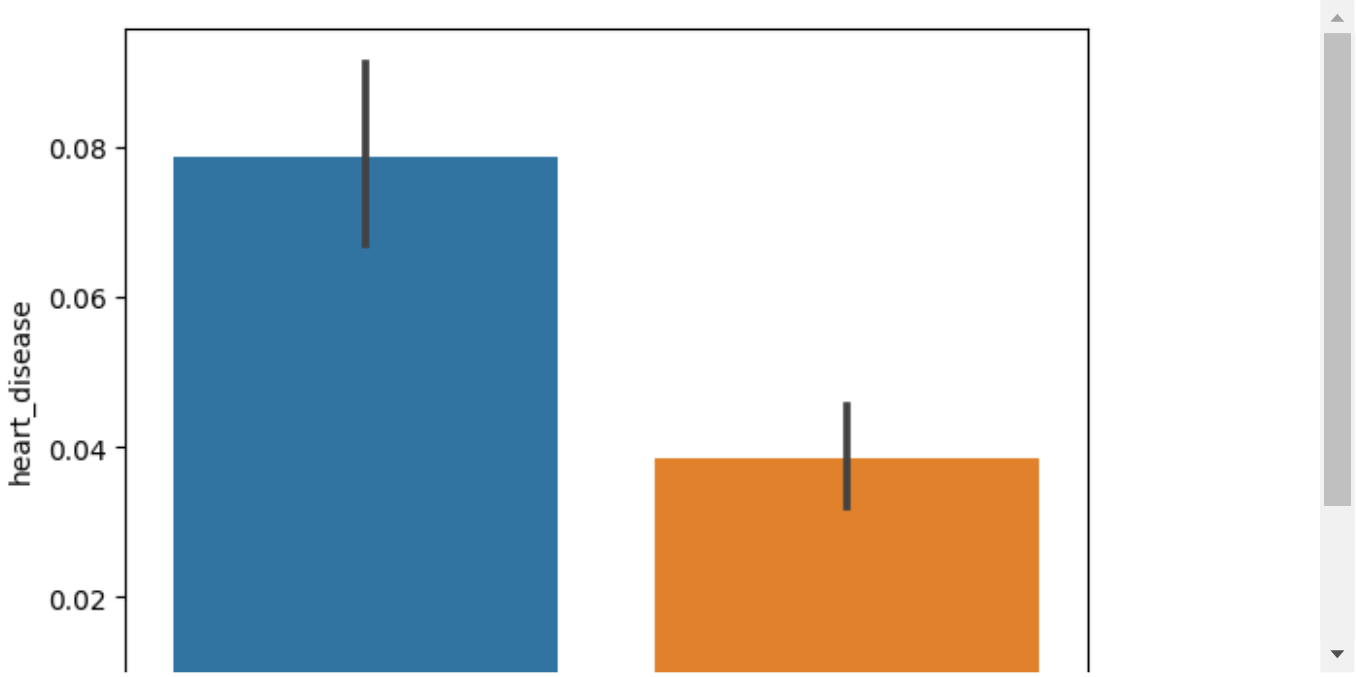
Построим график-диаграмму, где показан процент мужчин и женщин, болеющих когда-то сердечными заболеваниями.

In [14]:

```
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv('brain_stroke.csv')
df.dropna(inplace=True)

sns.barplot(data=df, x="gender", y="heart_disease")
plt.show()
```



Важно отметить, что в таблице представлены как люди, болеющие когда-либо ("1" в столбце "heart_disease"), так и никогда не болеющие соответствующими заболеваниями ("0" в столбце "heart_disease").

Исключаем второй случай.

In [29]:

```
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv('brain_stroke.csv')
df.dropna(inplace=True)
df.drop(df.index[df['heart_disease'] == 0], inplace=True)
# Оставляем только тех людей, которые болели.

df.head(20)
```

Out[29]:

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level
0	Male	67.0	0	1	Yes	Private	Urban	228.69
1	Male	80.0	0	1	Yes	Private	Rural	105.92
5	Male	74.0	1	1	Yes	Private	Rural	70.09
9	Female	61.0	0	1	Yes	Govt_job	Rural	120.46
11	Female	79.0	0	1	Yes	Private	Urban	214.09
13	Male	64.0	0	1	Yes	Private	Urban	191.61
19	Male	82.0	0	1	Yes	Private	Rural	208.30
23	Male	69.0	0	1	Yes	Self-employed	Urban	195.23
27	Male	80.0	0	1	Yes	Self-employed	Urban	252.72

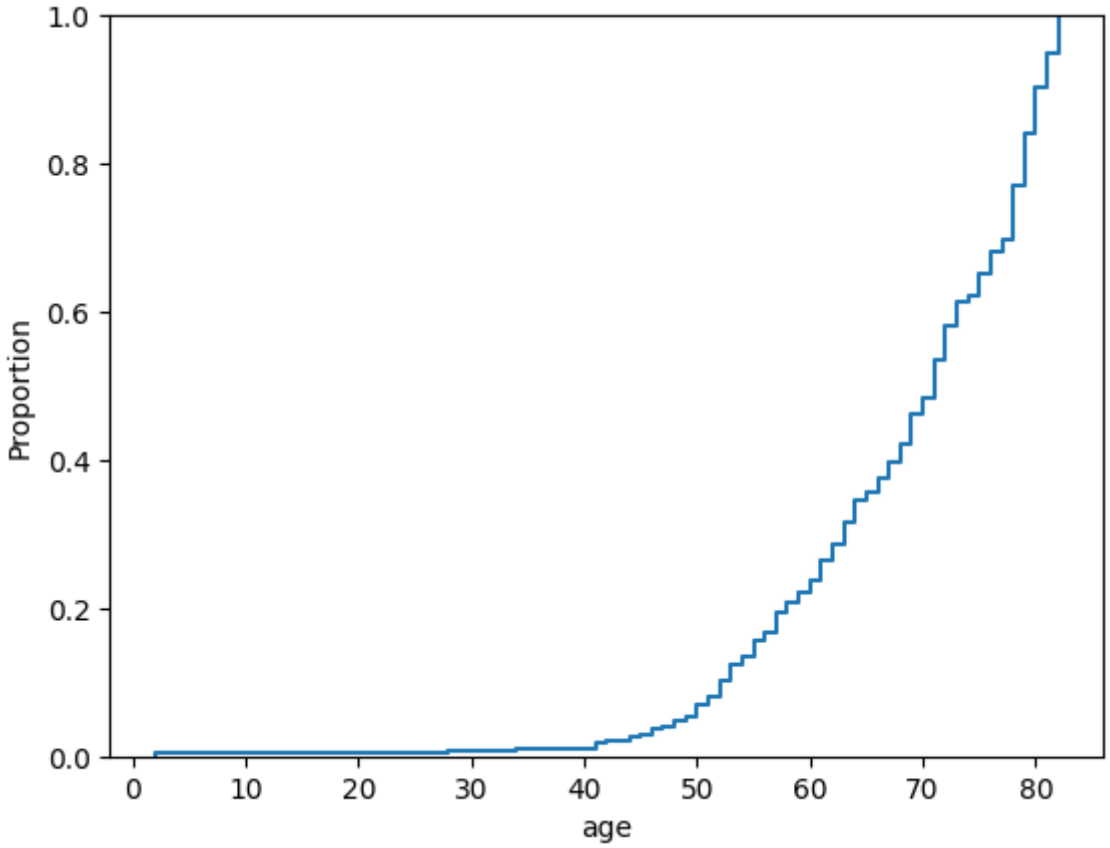
Изобразим график зависимости заболеваемости от возраста.

In [40]:

```
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

df = pd.read_csv('brain_stroke.csv')
df.dropna(inplace=True)
df.drop(df.index[df['heart_disease'] == 0], inplace=True)

sns.ecdfplot(data=df, x="age")
plt.show()
```



In [ ]:

```

```