

```
In [4]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [10]: d=pd.read_csv("C:\\Users\\samad\\Desktop\\dataset\\fulldata.csv")
df=pd.DataFrame(d)
print(d)
```

	gender	age	hypertension	heart_disease	ever_married	work_type \
0	Male	67.0	0	1	Yes	Private
1	Male	80.0	0	1	Yes	Private
2	Female	49.0	0	0	Yes	Private
3	Female	79.0	1	0	Yes	Self-employed
4	Male	81.0	0	0	Yes	Private
...
4976	Male	41.0	0	0	No	Private
4977	Male	40.0	0	0	Yes	Private
4978	Female	45.0	1	0	Yes	Govt_job
4979	Male	40.0	0	0	Yes	Private
4980	Female	80.0	1	0	Yes	Private

	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	Urban	228.69	36.6	formerly smoked	1
1	Rural	105.92	32.5	never smoked	1
2	Urban	171.23	34.4	smokes	1
3	Rural	174.12	24.0	never smoked	1
4	Urban	186.21	29.0	formerly smoked	1
...
4976	Rural	70.15	29.8	formerly smoked	0
4977	Urban	191.15	31.1	smokes	0
4978	Rural	95.02	31.8	smokes	0
4979	Rural	83.94	30.0	smokes	0
4980	Urban	83.75	29.1	never smoked	0

[4981 rows x 11 columns]

```
In [11]: d.head()
```

```
Out[11]:
```

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_
0	Male	67.0	0	1	Yes	Private	Urban	228.69
1	Male	80.0	0	1	Yes	Private	Rural	105.92
2	Female	49.0	0	0	Yes	Private	Urban	171.23
3	Female	79.0	1	0	Yes	Self-employed	Rural	174.12
4	Male	81.0	0	0	Yes	Private	Urban	186.21

```
In [12]: d.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4981 entries, 0 to 4980
Data columns (total 11 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   gender                 4981 non-null   object
1   age                   4981 non-null   float64
2   hypertension           4981 non-null   int64
3   heart_disease          4981 non-null   int64
4   ever_married           4981 non-null   object
5   work_type              4981 non-null   object
6   Residence_type         4981 non-null   object
7   avg_glucose_level      4981 non-null   float64
8   bmi                   4981 non-null   float64
9   smoking_status         4981 non-null   object
10  stroke                 4981 non-null   int64
dtypes: float64(3), int64(3), object(5)
memory usage: 428.2+ KB
```

In [14]: `d.tail()`

```
Out[14]:
```

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level
4976	Male	41.0	0	0	No	Private	Rural	
4977	Male	40.0	0	0	Yes	Private	Urban	
4978	Female	45.0	1	0	Yes	Govt_job	Rural	
4979	Male	40.0	0	0	Yes	Private	Rural	
4980	Female	80.0	1	0	Yes	Private	Urban	

In [13]: `d.columns`

```
Out[13]: Index(['gender', 'age', 'hypertension', 'heart_disease', 'ever_married',
              'work_type', 'Residence_type', 'avg_glucose_level', 'bmi',
              'smoking_status', 'stroke'],
              dtype='object')
```

In [17]: `d.isnull().sum()`

```
Out[17]: gender                0
age                0
hypertension       0
heart_disease      0
ever_married       0
work_type          0
Residence_type     0
avg_glucose_level  0
bmi                0
smoking_status     0
stroke             0
dtype: int64
```

In [19]: `d.describe()`

Out[19]:

	age	hypertension	heart_disease	avg_glucose_level	bmi	stroke
count	4981.000000	4981.000000	4981.000000	4981.000000	4981.000000	4981.000000
mean	43.419859	0.096165	0.055210	105.943562	28.498173	0.049789
std	22.662755	0.294848	0.228412	45.075373	6.790464	0.217531
min	0.080000	0.000000	0.000000	55.120000	14.000000	0.000000
25%	25.000000	0.000000	0.000000	77.230000	23.700000	0.000000
50%	45.000000	0.000000	0.000000	91.850000	28.100000	0.000000
75%	61.000000	0.000000	0.000000	113.860000	32.600000	0.000000
max	82.000000	1.000000	1.000000	271.740000	48.900000	1.000000

In [27]: `df.shape`

Out[27]: (4981, 11)

In [16]: `d.stroke.value_counts()`

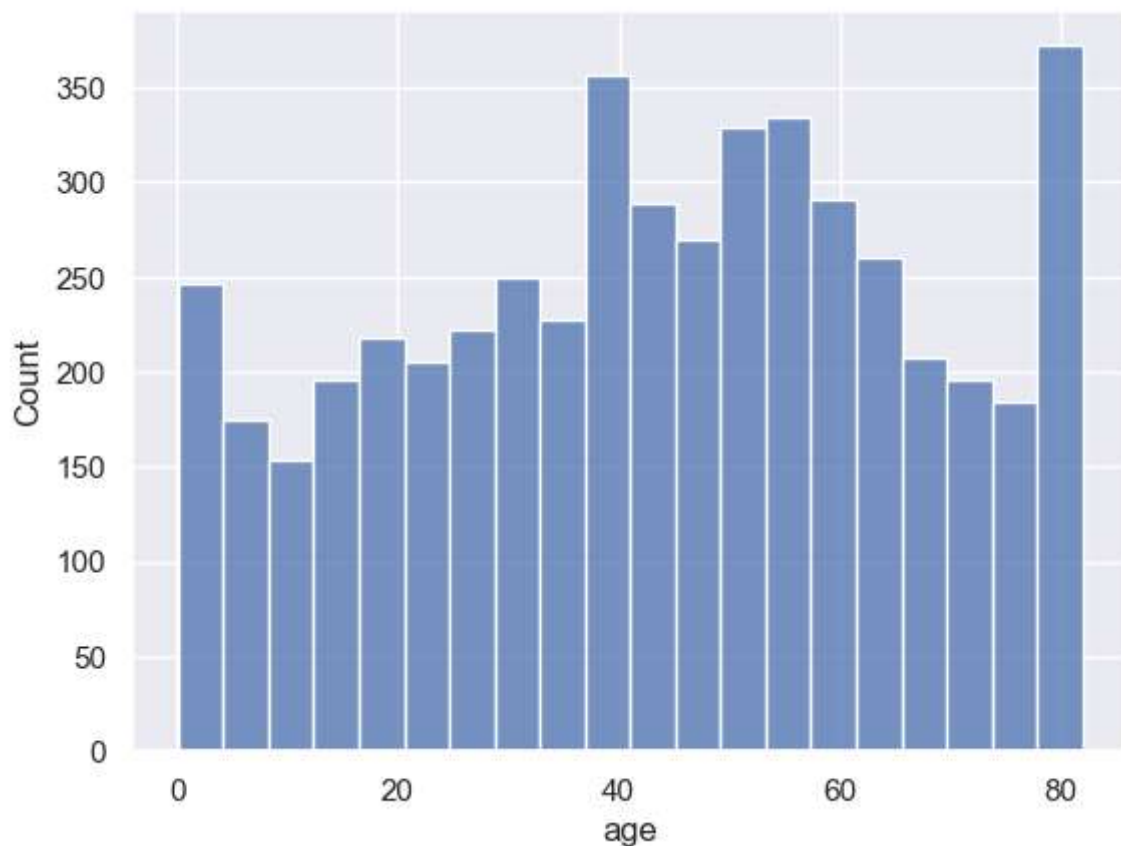
Out[16]:

```
0    4733
1     248
Name: stroke, dtype: int64
```

In [23]:

```
sns.set_theme(style='darkgrid')
sns.histplot(d.age)
```

Out[23]: <AxesSubplot:xlabel='age', ylabel='Count'>

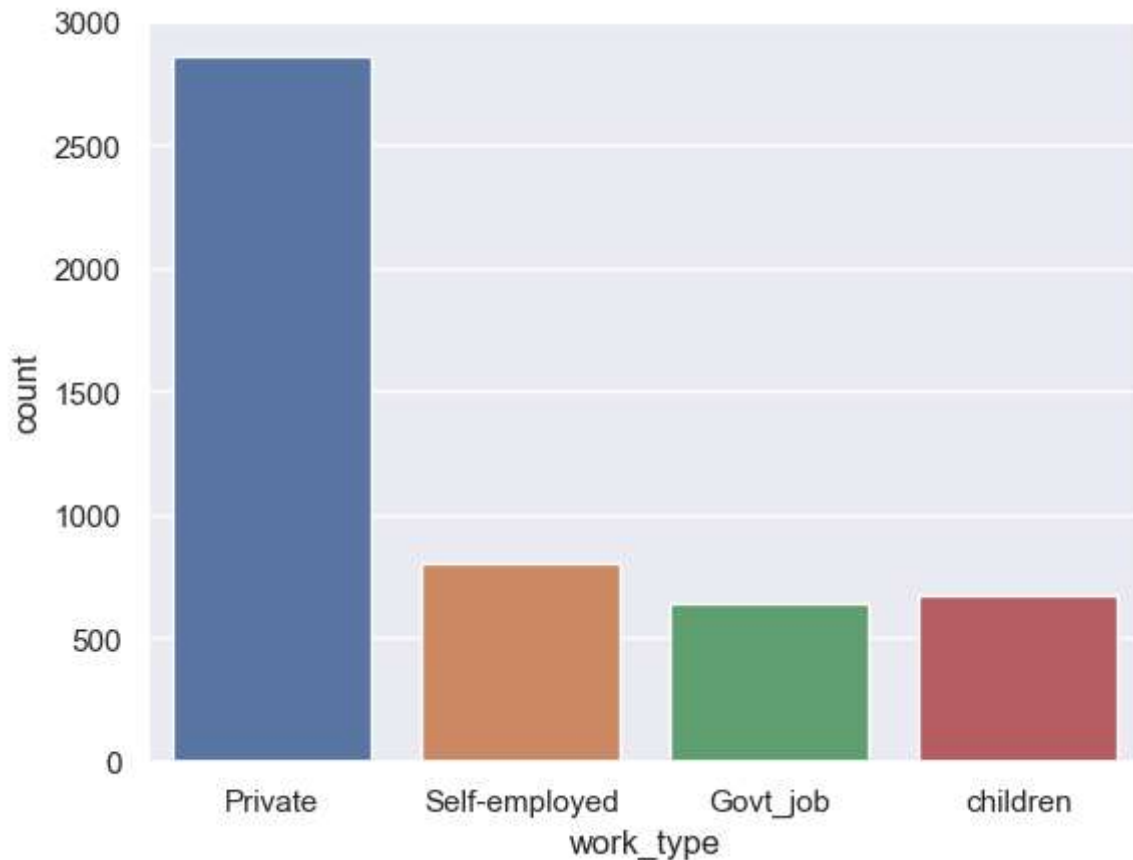


```
In [26]: sns.countplot(d.iloc[:,5])
```

C:\Program Files\Python310\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(
```

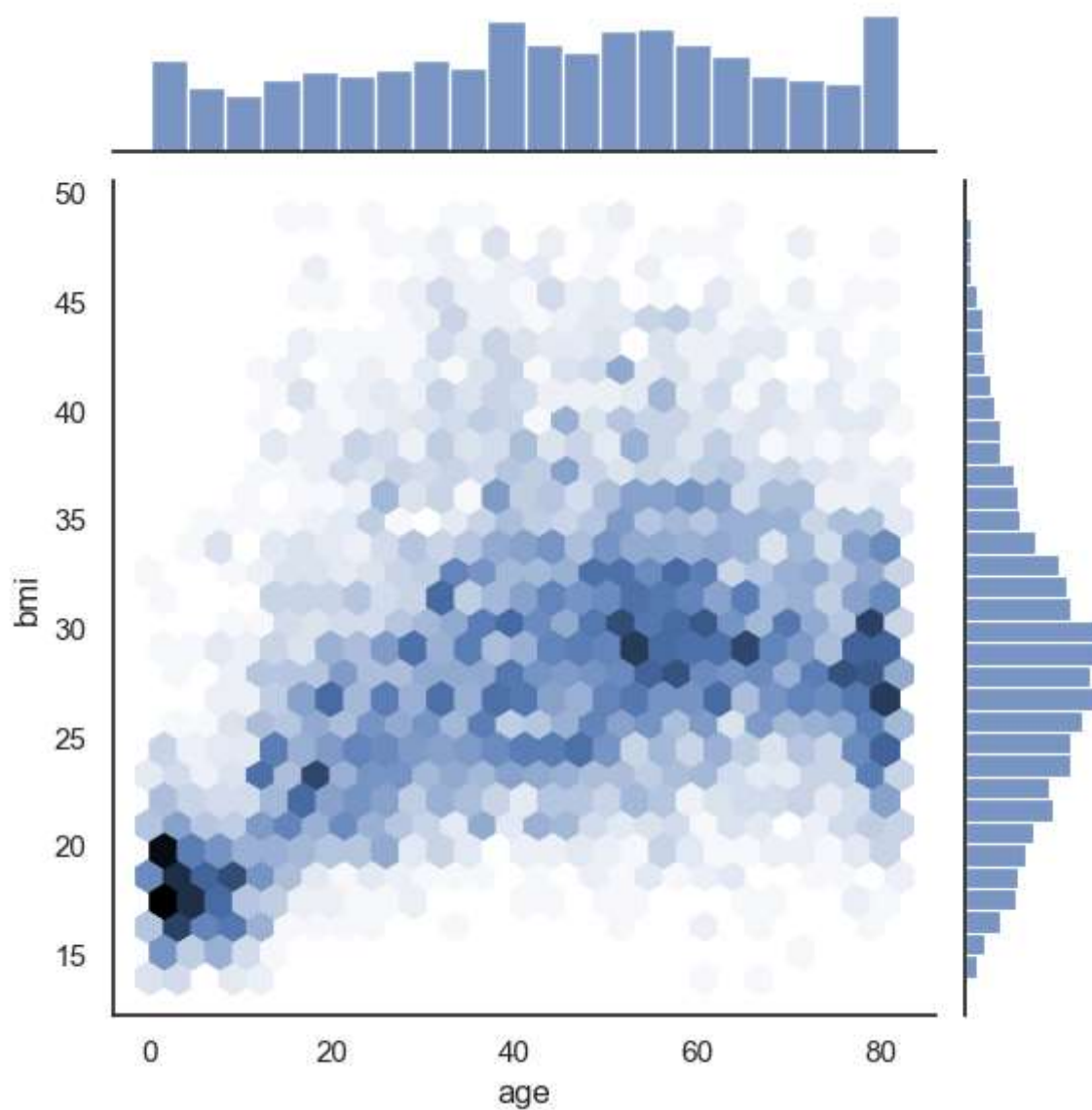
```
Out[26]: <AxesSubplot:xlabel='work_type', ylabel='count'>
```



```
In [28]: with sns.axes_style('white'):  
         sns.jointplot("age", "bmi" , d, kind='hex')
```

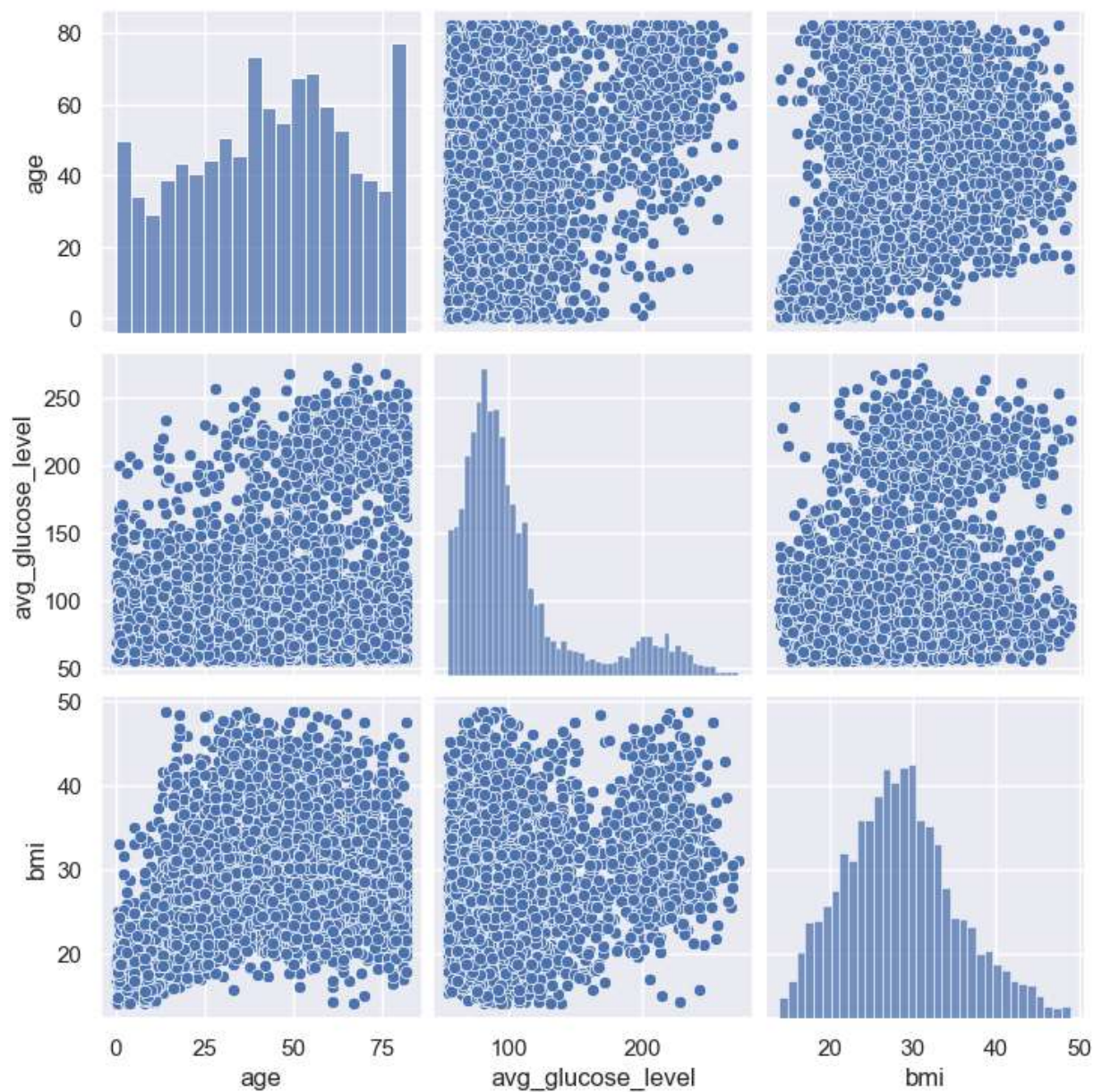
C:\Program Files\Python310\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y, data. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(
```



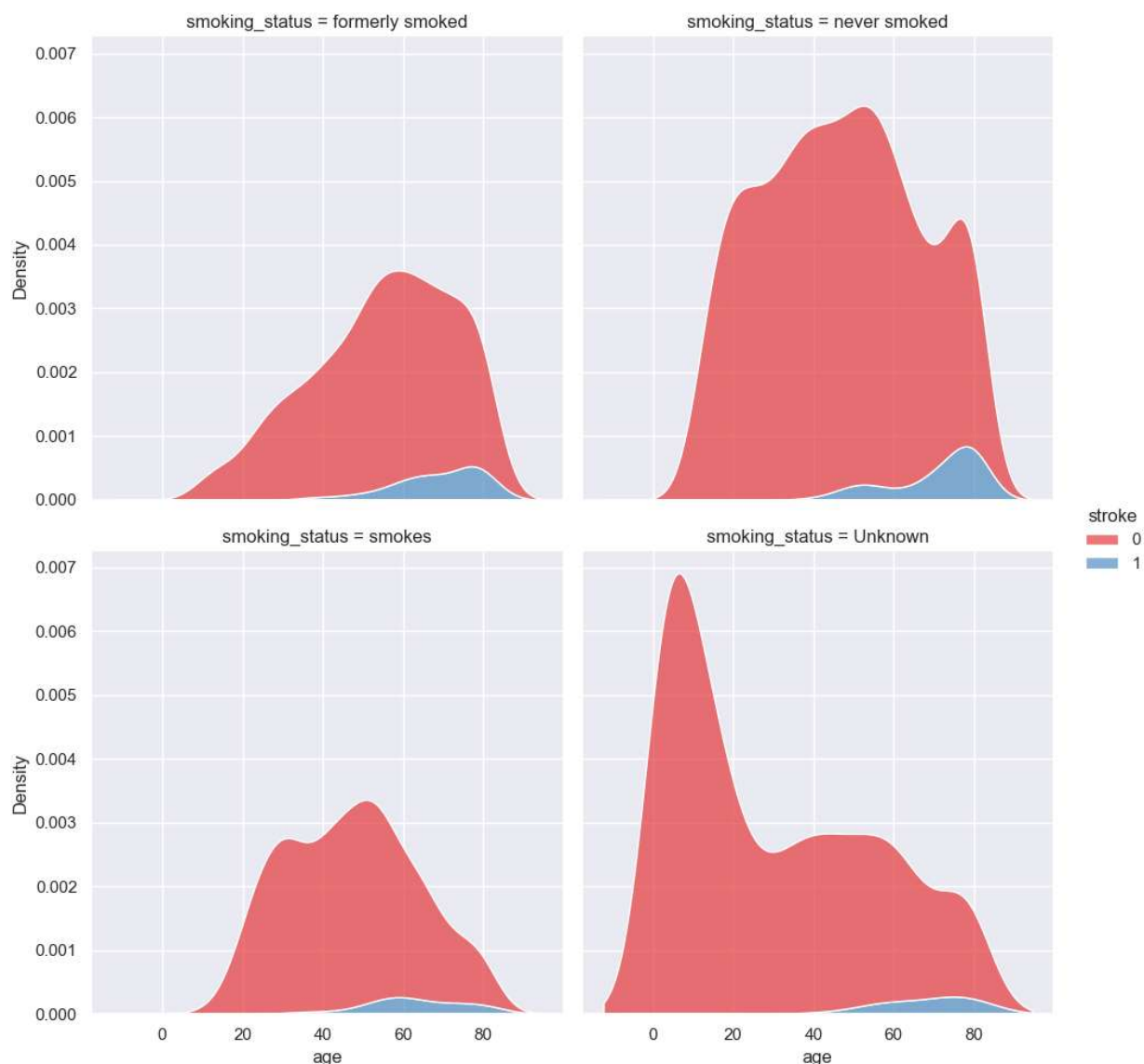
```
In [29]: sns.pairplot(df.loc[:, ['age', 'avg_glucose_level', 'bmi']], palette = 'Set1')
```

```
Out[29]: <seaborn.axisgrid.PairGrid at 0x1d84fc13850>
```

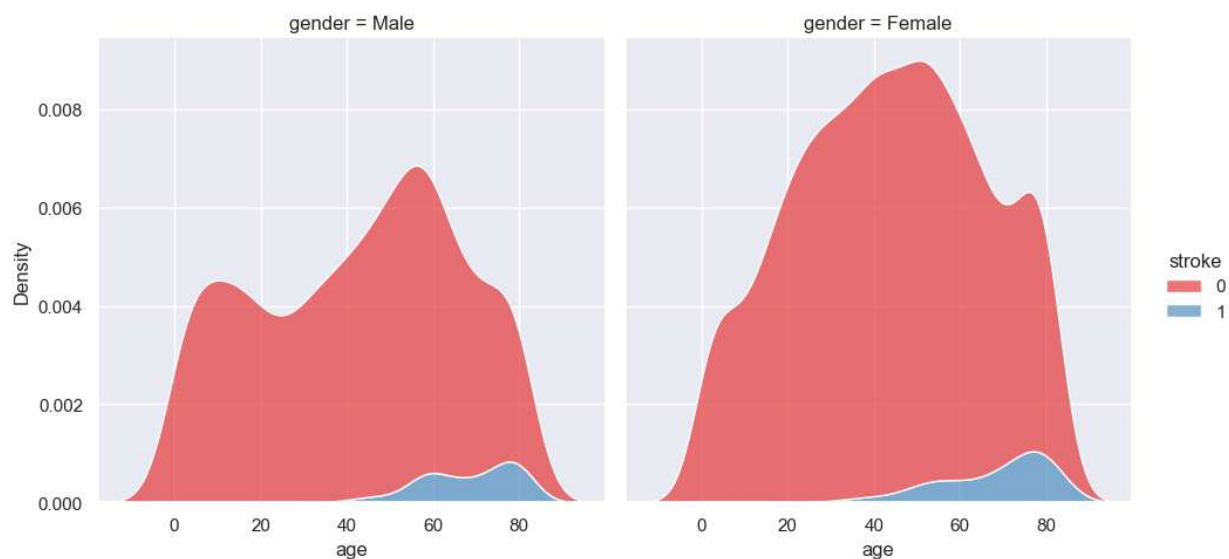
```
In [30]: sns.displot( df, x = 'age', hue = 'stroke', col = 'smoking_status', kind = 'kde',
                    multiple="stack", palette = 'Set1', alpha = 0.6, col_wrap = 2)
```

```
Out[30]: <seaborn.axisgrid.FacetGrid at 0x1d84fc13550>
```



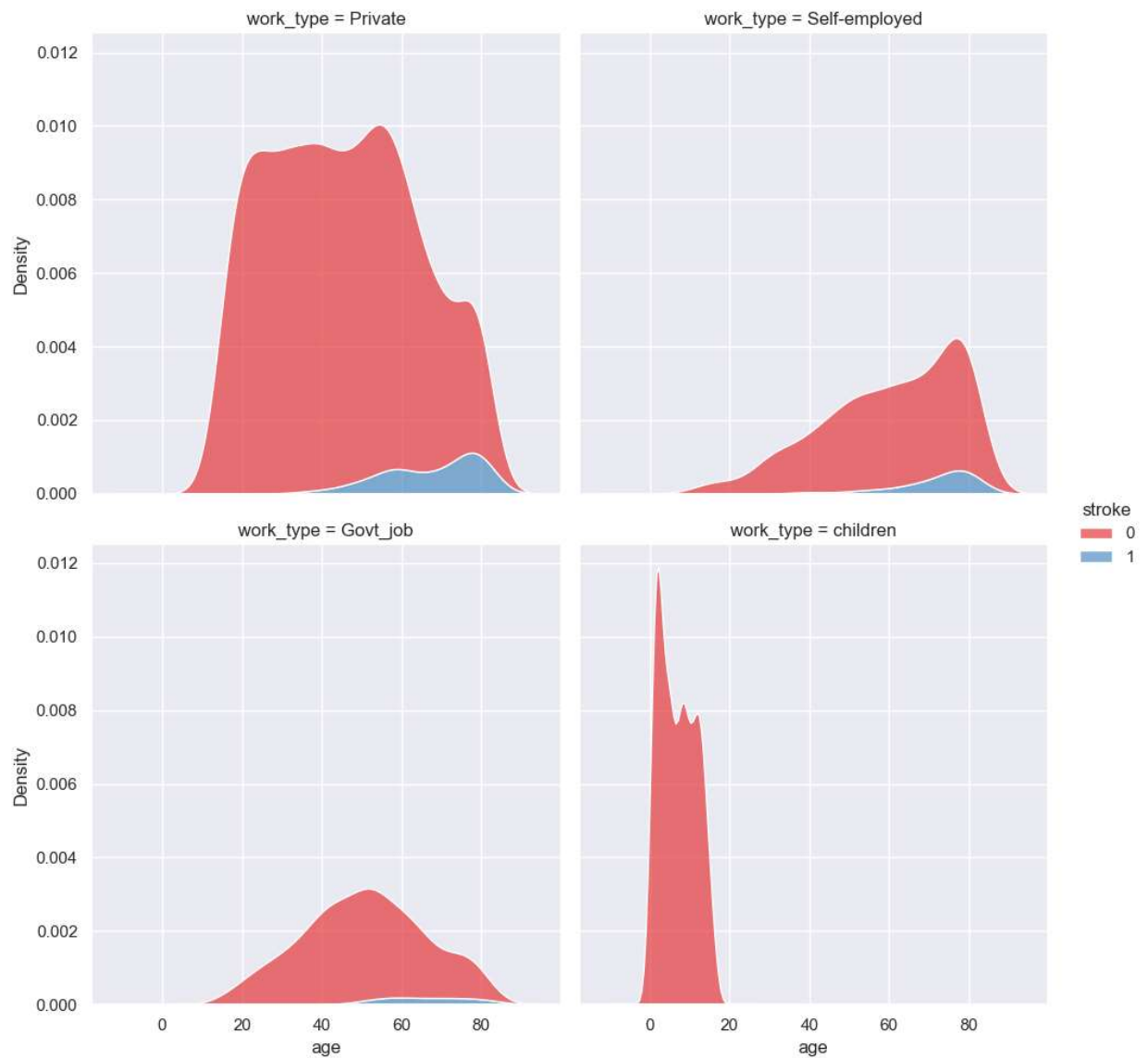
```
In [31]: sns.displot( df, x = 'age', hue = 'stroke', col = 'gender', multiple="stack", palette =
```

```
Out[31]: <seaborn.axisgrid.FacetGrid at 0x1d84fc131f0>
```



```
In [32]: sns.displot( df, x = 'age', hue = 'stroke', col = 'work_type', multiple="stack", palette =
```

Out[32]: <seaborn.axisgrid.FacetGrid at 0x1d850101960>



In []: