```
# Importing Libraries
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
plt.style.use('default')
```

```
df =pd.read_csv("data.csv")
```

```
df.head()
```

```
      id  gender   age  hypertension  heart_disease ever_married  \
0   9046    Male  67.0             0              1          Yes
1  51676  Female  61.0             0              0          Yes
2  31112    Male  80.0             0              1          Yes
3  60182  Female  49.0             0              0          Yes
4   1665  Female  79.0             1              0          Yes

       work_type Residence_type  avg_glucose_level   bmi
smoking_status  \
0        Private          Urban             228.69  36.6  formerly
smoked
1  Self-employed          Rural             202.21   NaN     never
smoked
2        Private          Rural             105.92  32.5     never
smoked
3        Private          Urban             171.23  34.4
smokes
4  Self-employed          Rural             174.12  24.0     never
smoked

   stroke
0       1
1       1
2       1
3       1
4       1
```

```
print("THE ROWS AND COLUMS OF DATA FRAME")
df.shape
```

```
THE ROWS AND COLUMS OF DATA FRAME
```

```
(5110, 12)
```

```
print("THE BASIC INFORMATION ABOUT DATA FRAME")
print(df.info())
```

```
THE BASIC INFORMATION ABOUT DATA FRAME
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 5110 entries, 0 to 5109
Data columns (total 12 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   id                 5110 non-null   int64  
 1   gender             5110 non-null   object 
 2   age                5110 non-null   float64
 3   hypertension       5110 non-null   int64  
 4   heart_disease      5110 non-null   int64  
 5   ever_married       5110 non-null   object 
 6   work_type          5110 non-null   object 
 7   Residence_type     5110 non-null   object 
 8   avg_glucose_level  5110 non-null   float64
 9   bmi                4909 non-null   float64
 10  smoking_status     5110 non-null   object 
 11  stroke             5110 non-null   int64  
dtypes: float64(3), int64(4), object(5)
memory usage: 479.2+ KB
None
```

```
print("TOTAL NO.OF UNIQUE VAULES")
print(df.id.unique())
```

```
TOTAL NO.OF UNIQUE VAULES
[ 9046 51676 31112 ... 19723 37544 44679]
```

```
print("SEARCHING FOR DUPLICATE VALUES")
df.duplicated().sum()
```

```
SEARCHING FOR DUPLICATE VALUES
```

```
np.int64(0)
```

```
df.describe()
```

```
                 id          age  hypertension  heart_disease  \
count   5110.000000  5110.000000   5110.000000    5110.000000   
mean   36517.829354    43.226614      0.097456       0.054012   
std    21161.721625    22.612647      0.296607       0.226063   
min       67.000000     0.080000      0.000000       0.000000   
25%    17741.250000    25.000000      0.000000       0.000000   
50%    36932.000000    45.000000      0.000000       0.000000   
75%    54682.000000    61.000000      0.000000       0.000000   
max    72940.000000    82.000000      1.000000       1.000000   

       avg_glucose_level          bmi       stroke  
count        5110.000000  4909.000000  5110.000000  
mean          106.147677    28.893237     0.048728  
std            45.283560     7.854067     0.215320  
min            55.120000    10.300000     0.000000  
25%            77.245000    23.500000     0.000000  
```

```
50%             91.885000    28.100000     0.000000
75%            114.090000    33.100000     0.000000
max            271.740000    97.600000     1.000000
```

```
print("TOTAL NO.OF NULL VALUES")
print(df.isnull().sum())
```

```
TOTAL NO.OF NULL VALUES
id                       0
gender                   0
age                      0
hypertension             0
heart_disease            0
ever_married             0
work_type                0
Residence_type           0
avg_glucose_level        0
bmi                    201
smoking_status           0
stroke                   0
dtype: int64
```

```
print(df.groupby('gender')['stroke'].mean())
```

```
gender
Female    0.047094
Male      0.051064
Other     0.000000
Name: stroke, dtype: float64
```

```
print(df.groupby('work_type')['stroke'].mean())
```

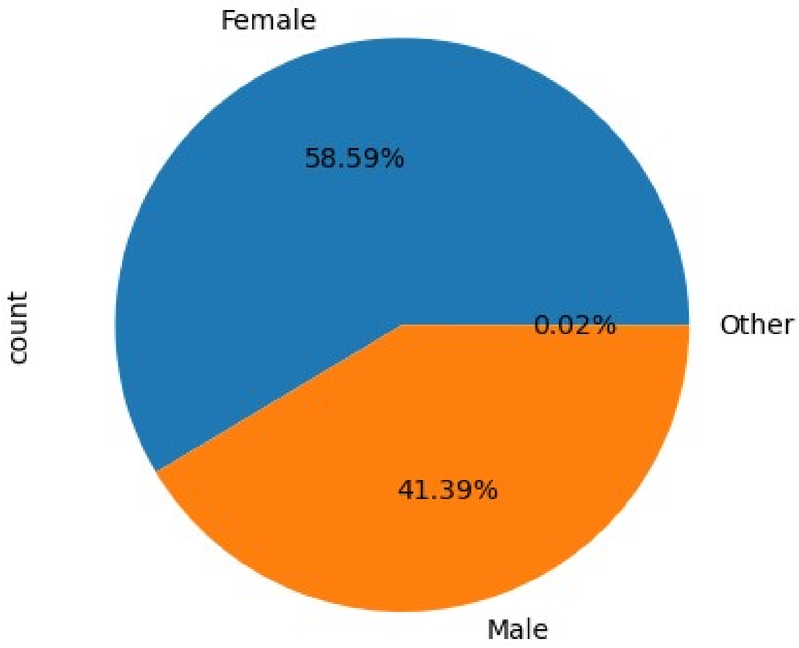
```
work_type
Govt_job         0.050228
Never_worked     0.000000
Private          0.050940
Self-employed    0.079365
children         0.002911
Name: stroke, dtype: float64
```

```
print(df.groupby('Residence_type')['stroke'].mean())
```

```
Residence_type
Rural    0.045346
Urban    0.052003
Name: stroke, dtype: float64
```

```
(5110, 12)
df["gender"].value_counts().plot(kind="pie",autopct='%1.2f%%')
```

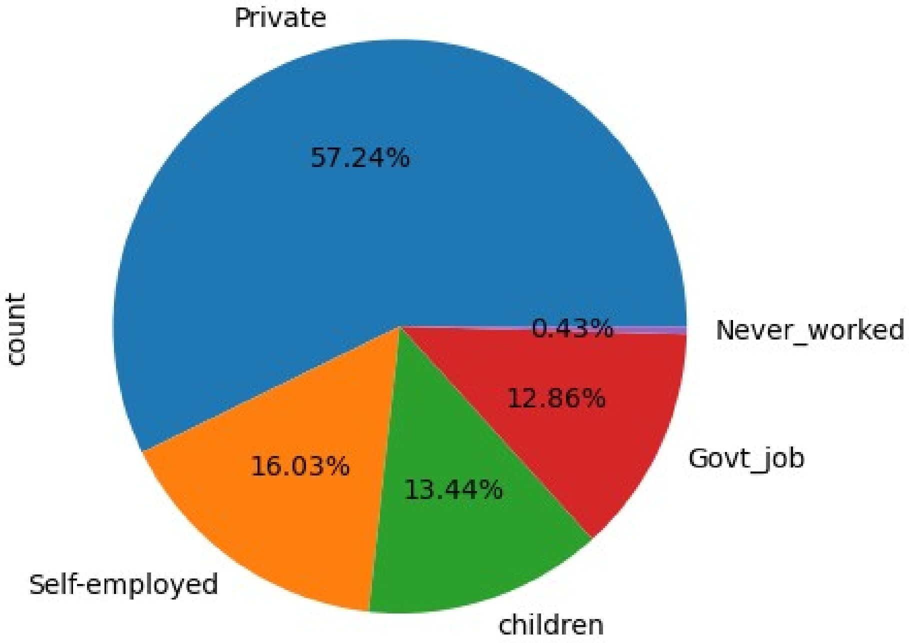
```
<Axes: ylabel='count'>
```



```
print("WORK TYPE OF PATIENTS")
df["work_type"].value_counts().plot(kind="pie",autopct='%1.2f%%')
```

```
WORK TYPE OF PATIENTS
```

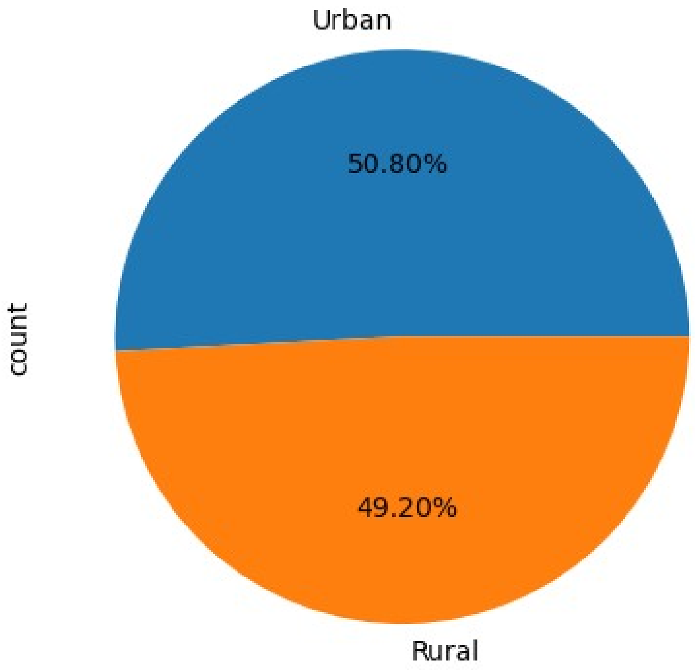
```
<Axes: ylabel='count'>
```



```
print("PATIENT'S RESIDENTIAL AREA ")
df["Residence_type"].value_counts().plot(kind="pie",autopct='%1.2f%%')
```

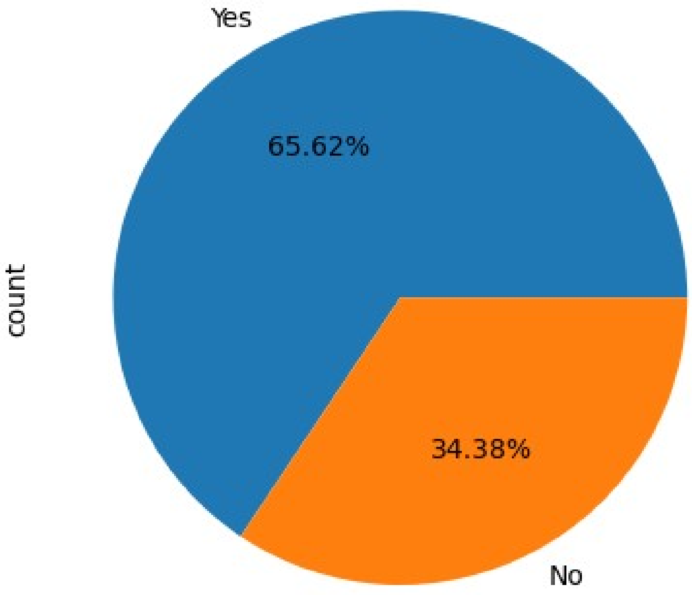
```
PATIENT'S RESIDENTIAL AREA
```

```
<Axes: ylabel='count'>
```



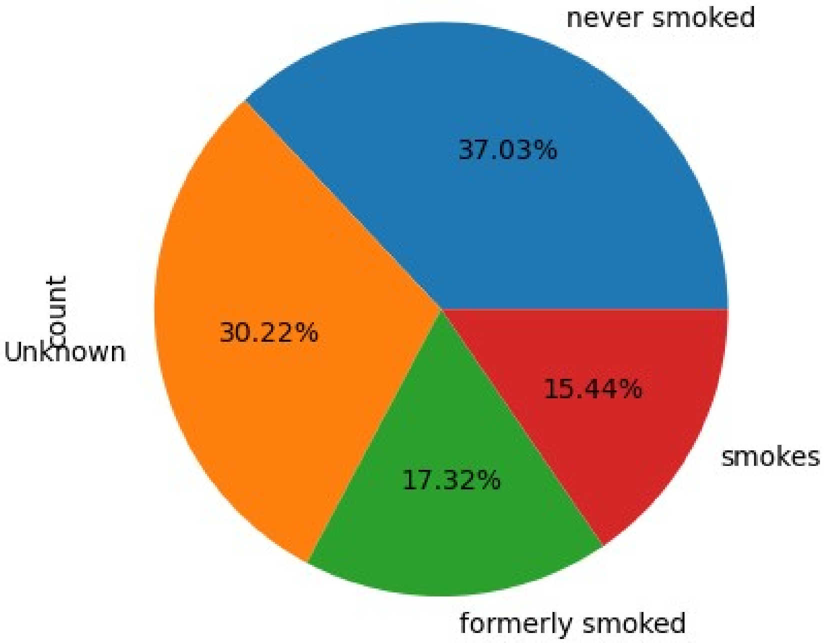
```
df.heart_disease.sum()
276
df.hypertension.sum()
498
df.heart_disease.sum()
276
df["ever_married"].value_counts().plot(kind="pie",autopct='%1.2f%%')

<Axes: ylabel='count'>
```



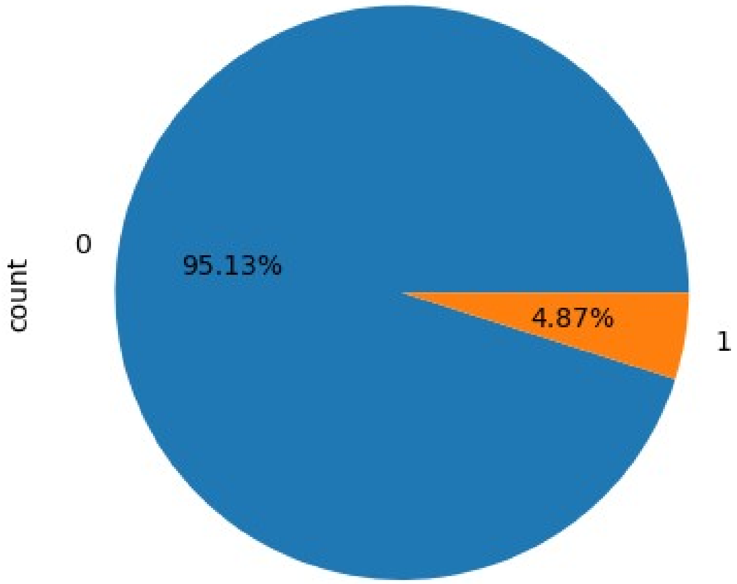
```
df["smoking_status"].value_counts().plot(kind="pie",autopct='%1.2f%%')
```

```
<Axes: ylabel='count'>
```



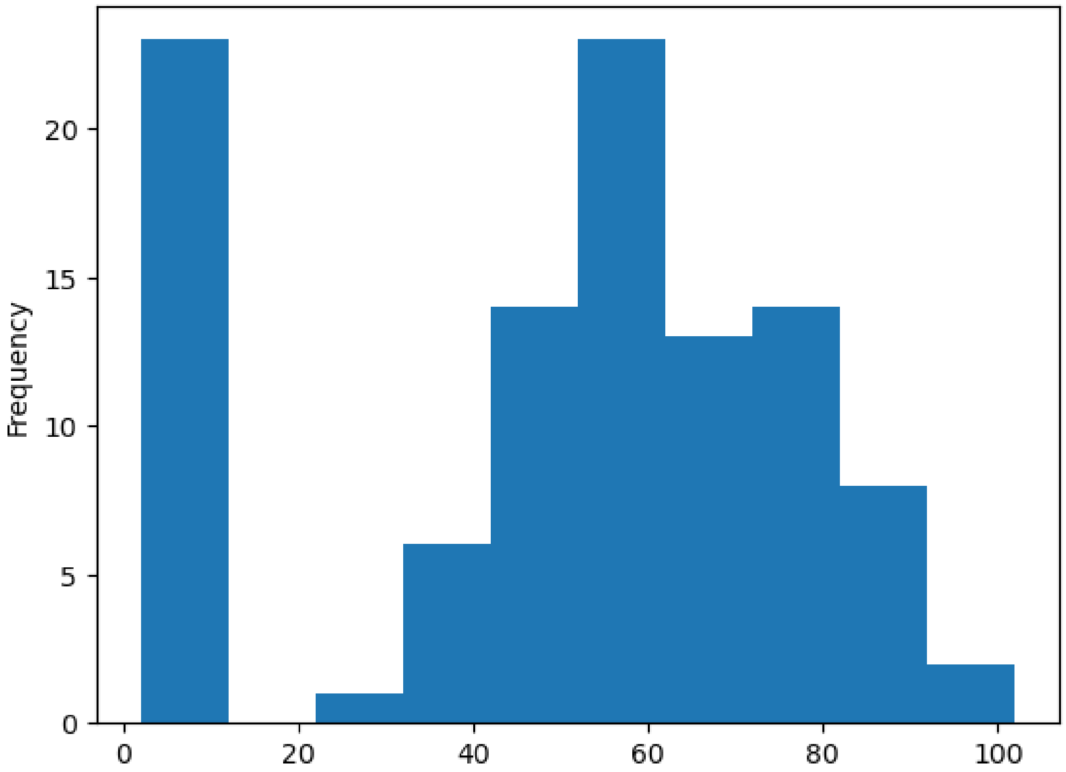
```
df["stroke"].value_counts().plot(kind="pie",autopct='%1.2f%%')
```

```
<Axes: ylabel='count'>
```



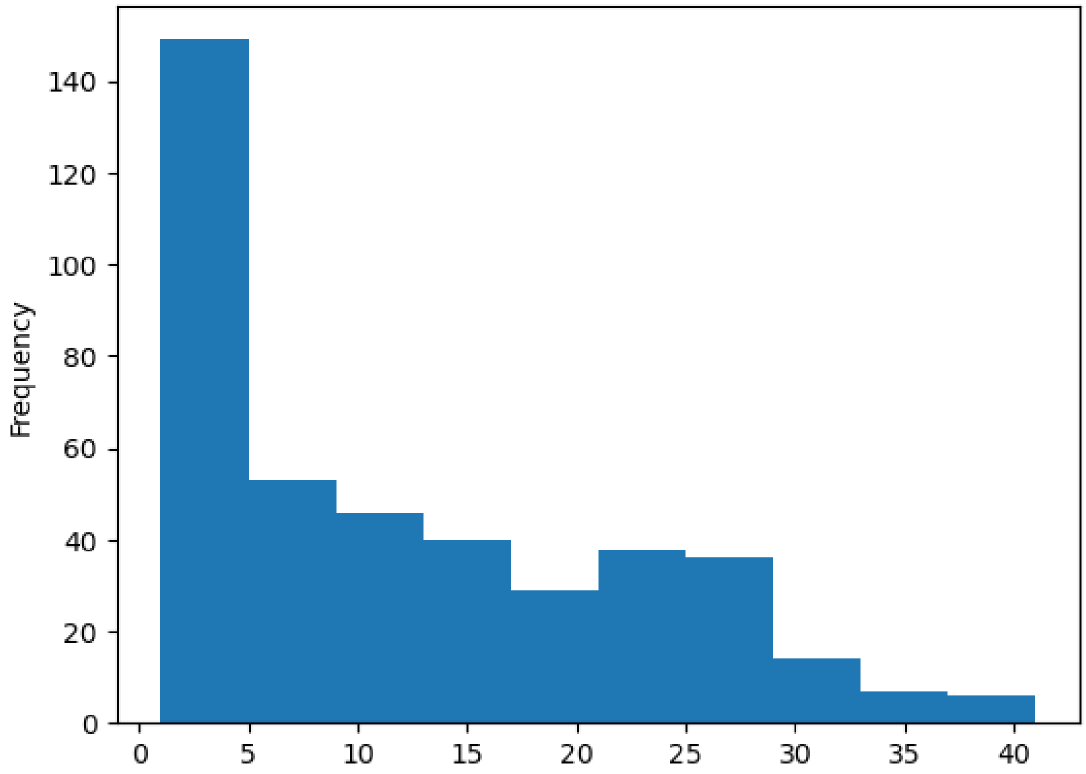
```
df["age"].value_counts().plot(kind="hist")
```

```
<Axes: ylabel='Frequency'>
```



```
df["bmi"].value_counts().plot(kind="hist")
```

```
<Axes: ylabel='Frequency'>
```



□ Analytical Report

After analyzing the dataset, I found that strokes mostly happen in older people (above 60 years). People who have high glucose levels and high BMI (overweight) are more likely to have a stroke. Also, those with hypertension and heart disease face higher risk.

From the graphs, I observed that:

Females have slightly higher stroke rates than males.

Urban people have more strokes than rural people.

“Formerly smoked” people are more in the stroke group.

In conclusion, age, glucose level, BMI, hypertension, and heart disease are the main factors linked with stroke. Living a healthy life, avoiding smoking, and keeping sugar and weight normal can help prevent strokes.