

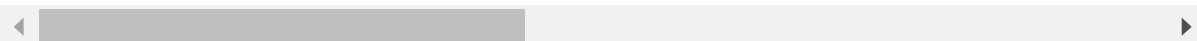
```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [12]: df=pd.read_csv(r"C:\Users\manasa\Downloads\archive (2).zip")
df
```

Out[12]:

	HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	D
0	No	16.60	Yes	No	No	3.0	30.0	
1	No	20.34	No	No	Yes	0.0	0.0	
2	No	26.58	Yes	No	No	20.0	30.0	
3	No	24.21	No	No	No	0.0	0.0	
4	No	23.71	No	No	No	28.0	0.0	
...
319790	Yes	27.41	Yes	No	No	7.0	0.0	
319791	No	29.84	Yes	No	No	0.0	0.0	
319792	No	24.24	No	No	No	0.0	0.0	
319793	No	32.81	No	No	No	0.0	0.0	
319794	No	46.56	No	No	No	0.0	0.0	

319795 rows × 18 columns



In [13]: df.info

Out[13]: <bound method DataFrame.info of
 nking Stroke PhysicalHealth
 0 No 16.60 Yes No No 3.0 \

	Stroke	PhysicalHealth	HeartDisease	BMI	Smoking	AlcoholDri
0	No	16.60	Yes	No	No	3.0
1	No	20.34	No	No	Yes	0.0
2	No	26.58	Yes	No	No	20.0
3	No	24.21	No	No	No	0.0
4	No	23.71	No	No	No	28.0
...
319790	Yes	27.41	Yes	No	No	7.0
319791	No	29.84	Yes	No	No	0.0
319792	No	24.24	No	No	No	0.0
319793	No	32.81	No	No	No	0.0
319794	No	46.56	No	No	No	0.0

	MentalHealth	DiffWalking	Sex	AgeCategory	Race	Diabetic
0	30.0	No	Female	55-59	White	Yes
1	0.0	No	Female	80 or older	White	No
2	30.0	No	Male	65-69	White	Yes
3	0.0	No	Female	75-79	White	No
4	0.0	Yes	Female	40-44	White	No
...
319790	0.0	Yes	Male	60-64	Hispanic	Yes
319791	0.0	No	Male	35-39	Hispanic	No
319792	0.0	No	Female	45-49	Hispanic	No
319793	0.0	No	Female	25-29	Hispanic	No
319794	0.0	No	Female	80 or older	Hispanic	No

	PhysicalActivity	GenHealth	SleepTime	Asthma	KidneyDisease	SkinCancer
0	Yes	Very good	5.0	Yes	No	Yes
1	Yes	Very good	7.0	No	No	No
2	Yes	Fair	8.0	Yes	No	No
3	No	Good	6.0	No	No	Yes
4	Yes	Very good	8.0	No	No	No
...
319790	No	Fair	6.0	Yes	No	No
319791	Yes	Very good	5.0	Yes	No	No
319792	Yes	Good	6.0	No	No	No
319793	No	Good	12.0	No	No	No
319794	Yes	Good	8.0	No	No	No

[319795 rows x 18 columns]>

```
In [14]: df.isnull().sum()
```

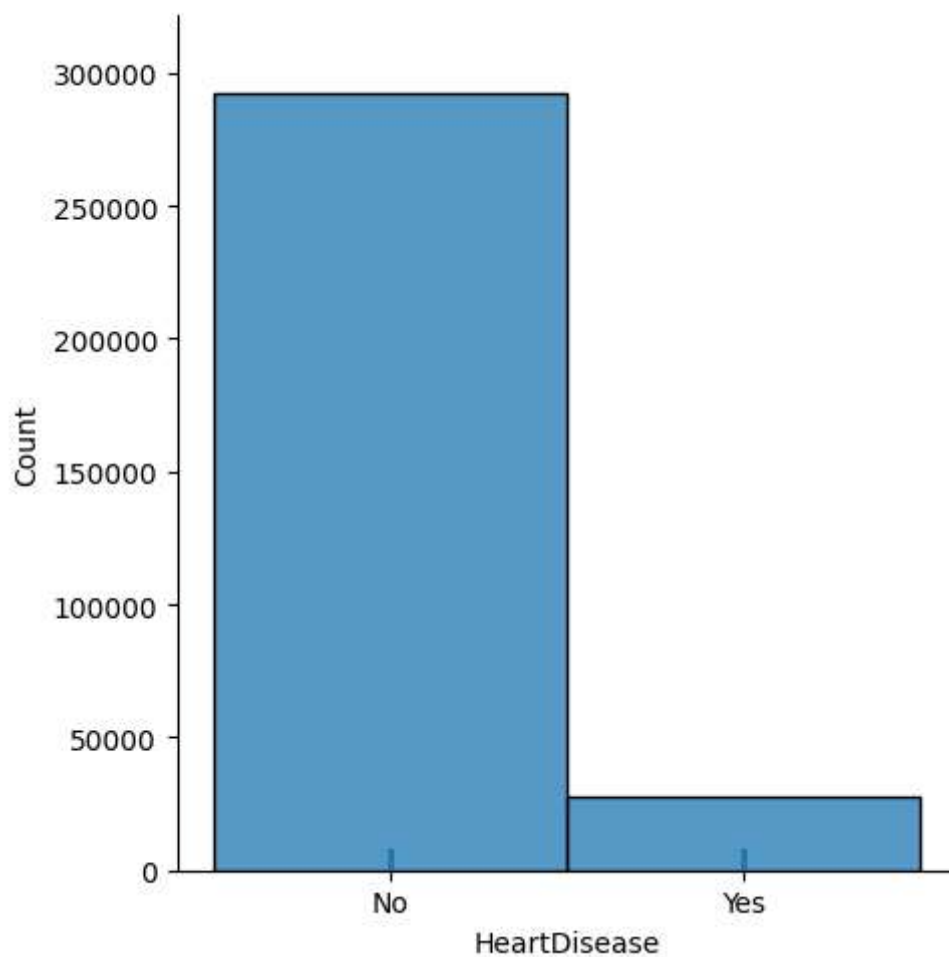
```
Out[14]: HeartDisease      0
        BMI                0
        Smoking            0
        AlcoholDrinking    0
        Stroke             0
        PhysicalHealth      0
        MentalHealth       0
        DiffWalking        0
        Sex                0
        AgeCategory        0
        Race               0
        Diabetic           0
        PhysicalActivity    0
        GenHealth          0
        SleepTime          0
        Asthma             0
        KidneyDisease       0
        SkinCancer         0
        dtype: int64
```

```
In [15]: df.describe()
```

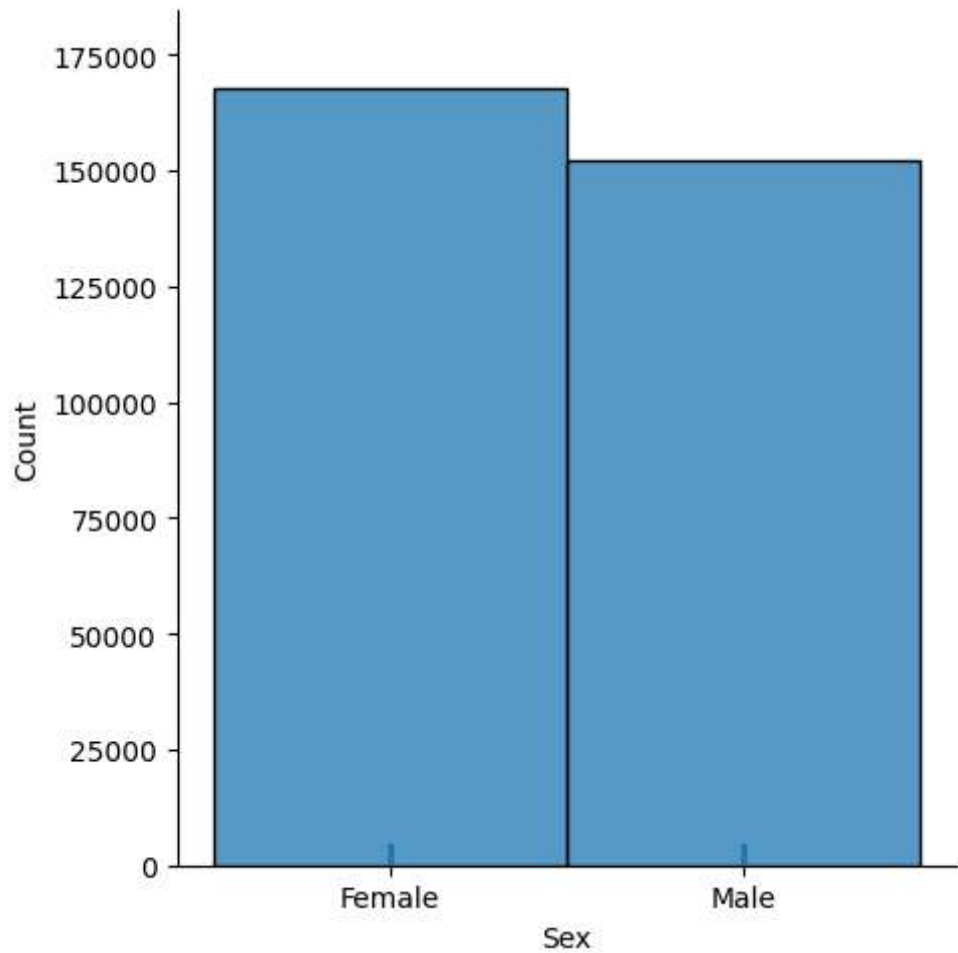
```
Out[15]:
```

	BMI	PhysicalHealth	MentalHealth	SleepTime
count	319795.000000	319795.000000	319795.000000	319795.000000
mean	28.325399	3.37171	3.898366	7.097075
std	6.356100	7.95085	7.955235	1.436007
min	12.020000	0.00000	0.000000	1.000000
25%	24.030000	0.00000	0.000000	6.000000
50%	27.340000	0.00000	0.000000	7.000000
75%	31.420000	2.00000	3.000000	8.000000
max	94.850000	30.00000	30.000000	24.000000

```
In [18]: import seaborn as sns  
sns.displot(df['HeartDisease'], rug=True)  
plt.show()
```



```
In [19]: sns.displot(df['Sex'],rug=True)  
plt.show()
```

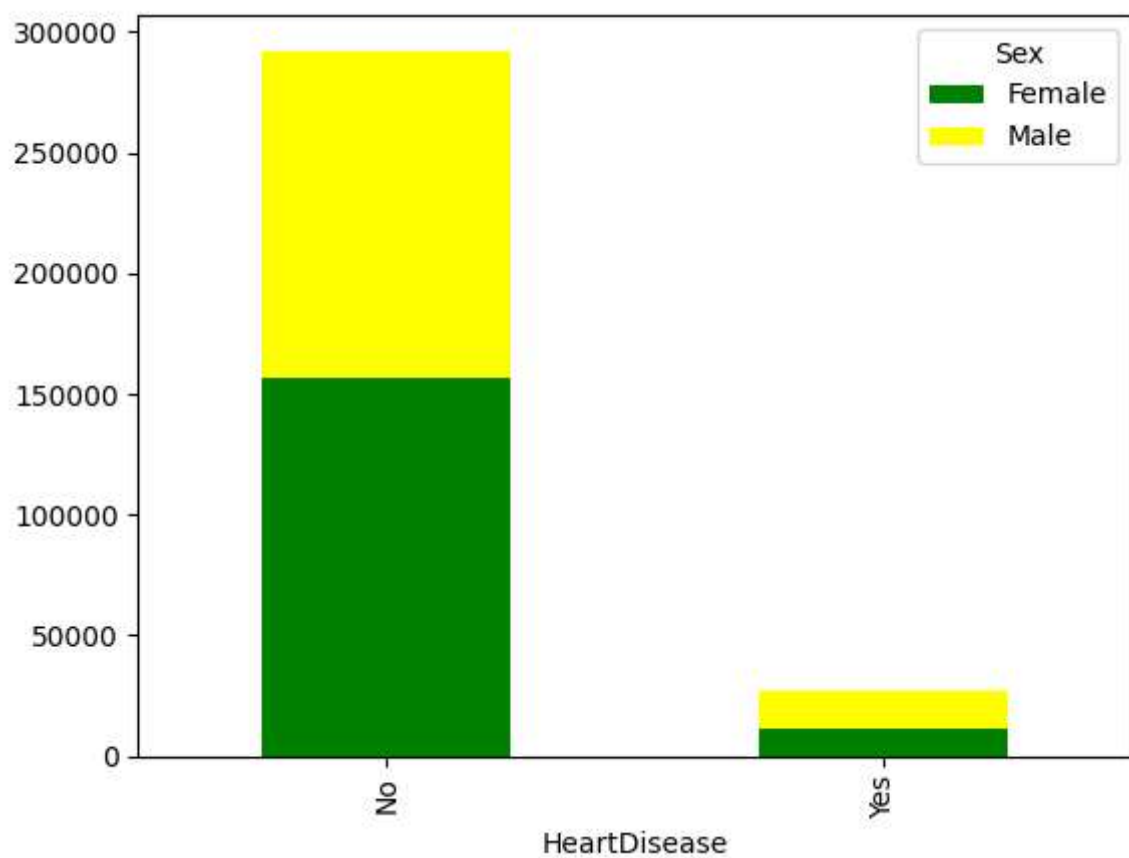


```
In [20]: s=pd.crosstab(df['HeartDisease'],df['Sex'])  
print(s)
```

Sex	Female	Male
HeartDisease		
No	156571	135851
Yes	11234	16139

```
In [21]: s.plot(kind='bar', stacked=True, color=['green','yellow'],grid=False)
```

```
Out[21]: <Axes: xlabel='HeartDisease'>
```



```
In [ ]:
```