

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

```
In [2]: df=pd.read_csv(r"C:\Users\chinta pavani\Downloads\archive (1).zip")
df
```

Out[2]:

	HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth
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 × 8 columns

In [3]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 319795 entries, 0 to 319794
Data columns (total 18 columns):
 #   Column                Non-Null Count  Dtype  
---  -
 0   HeartDisease          319795 non-null object  
 1   BMI                   319795 non-null float64
 2   Smoking               319795 non-null object  
 3   AlcoholDrinking       319795 non-null object  
 4   Stroke                319795 non-null object  
 5   PhysicalHealth         319795 non-null float64
 6   MentalHealth          319795 non-null float64
 7   DiffWalking           319795 non-null object  
 8   Sex                   319795 non-null object  
 9   AgeCategory           319795 non-null object  
10   Race                  319795 non-null object  
11   Diabetic              319795 non-null object  
12   PhysicalActivity       319795 non-null object  
13   GenHealth             319795 non-null object  
14   SleepTime             319795 non-null float64
15   Asthma                319795 non-null object  
16   KidneyDisease         319795 non-null object  
17   SkinCancer            319795 non-null object  
dtypes: float64(4), object(14)
memory usage: 43.9+ MB
```

In [4]: `df.isnull().sum()`

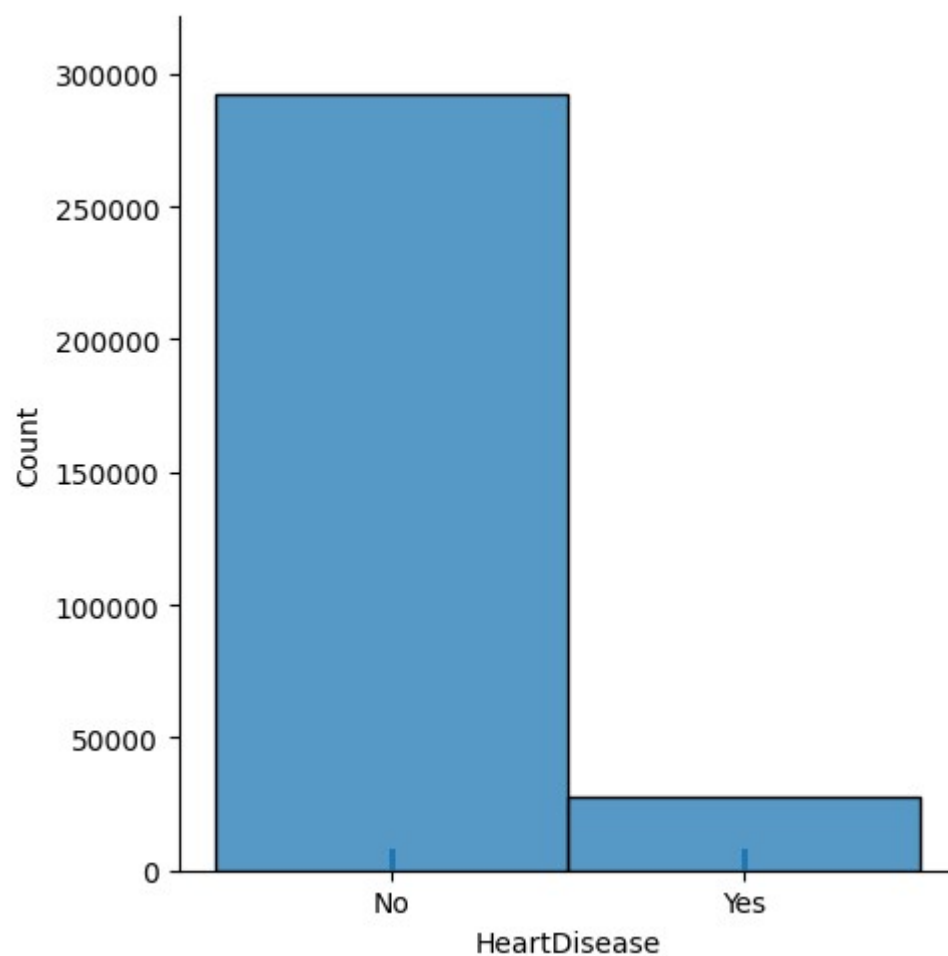
```
Out[4]: HeartDisease    0
        BMI             0
        Smoking         0
        AlcoholDrinking 0
        Stroke          0
        PhysicalHealth   0
        MentalHealth     0
        DiffWalking      0
        Sex              0
        AgeCategory      0
        Race             0
        Diabetic         0
        PhysicalActivity  0
        GenHealth        0
        SleepTime        0
        Asthma           0
        KidneyDisease    0
        SkinCancer       0
dtype: int64
```

```
In [5]: df.describe()
```

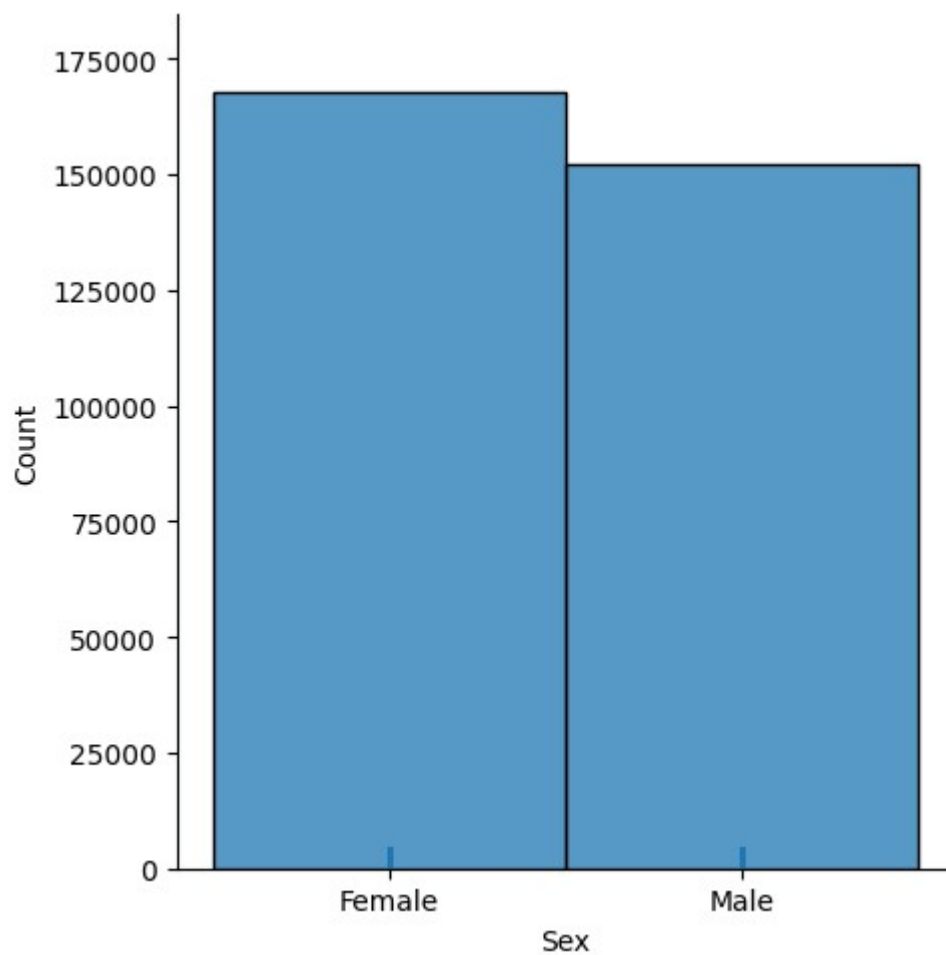
Out[5]:

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



```
In [8]: ▶ sns.displot(df['Sex'],rug=True)  
plt.show()
```

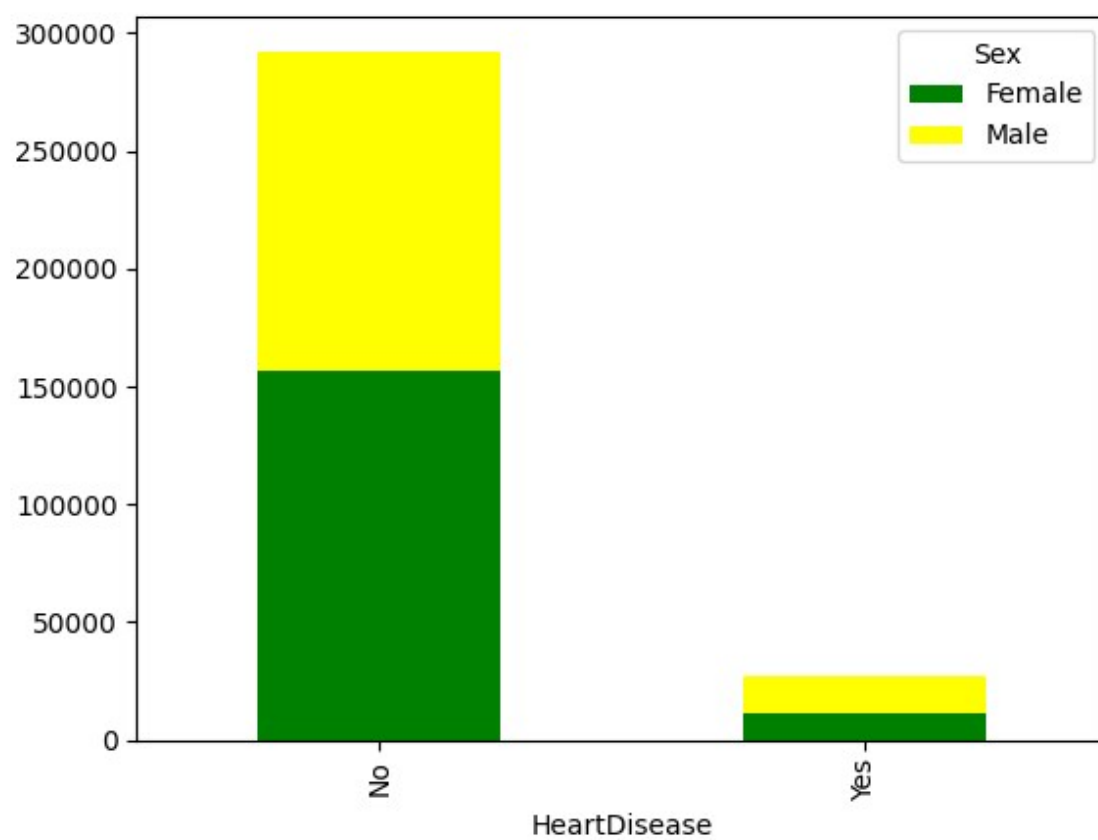


```
In [9]: ▶ s=pd.crosstab(df['HeartDisease'],df['Sex'])  
print(s)
```

Sex	Female	Male
HeartDisease		
No	156571	135851
Yes	11234	16139

```
In [10]: s.plot(kind='bar', stacked=True, color=['green','yellow'],grid=False)
```

```
Out[10]: <Axes: xlabel='HeartDisease'>
```



```
In [ ]: 
```