In [1]:
```
import numpy as np
import pandas as pd
from sklearn import preprocessing
import matplotlib.pyplot as plt
import seaborn as sns
sns.set(style="white")
sns.set(style="whitegrid", color_codes=True)
import warnings
warnings.simplefilter(action='ignore')
```

In [2]:
```
df=pd.read_csv(r"C:\Users\91756\Downloads\heart_2020_cleaned.csv.zip")
df
```

Out[2]:

	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 [3]:
```
df.head()
```

Out[3]:

	HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	DiffWal
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	

In [4]:
```
df.describe
```

Out[4]:
```
<bound method NDFrame.describe of        HeartDisease    BMI Smoking Alcohol
Drinking Stroke  PhysicalHealth
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 SkinCance
r
0                   Yes  Very good        5.0    Yes            No        Ye
s
1                   Yes  Very good        7.0     No            No         N
o
2                   Yes       Fair        8.0    Yes            No         N
o
3                    No       Good        6.0     No            No        Ye
s
4                   Yes  Very good        8.0     No            No         N
o
...                 ...        ...        ...    ...           ...
...
319790               No       Fair        6.0    Yes            No         N
o
319791              Yes  Very good        5.0    Yes            No         N
o
319792              Yes       Good        6.0     No            No         N
o
319793               No       Good       12.0     No            No         N
o
319794              Yes       Good        8.0     No            No         N
o

[319795 rows x 18 columns]>
```

In [5]:
```
df.shape
```

Out[5]:
```
(319795, 18)
```

In [6]:
```
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 [7]:
```
df.isnull().sum()
```

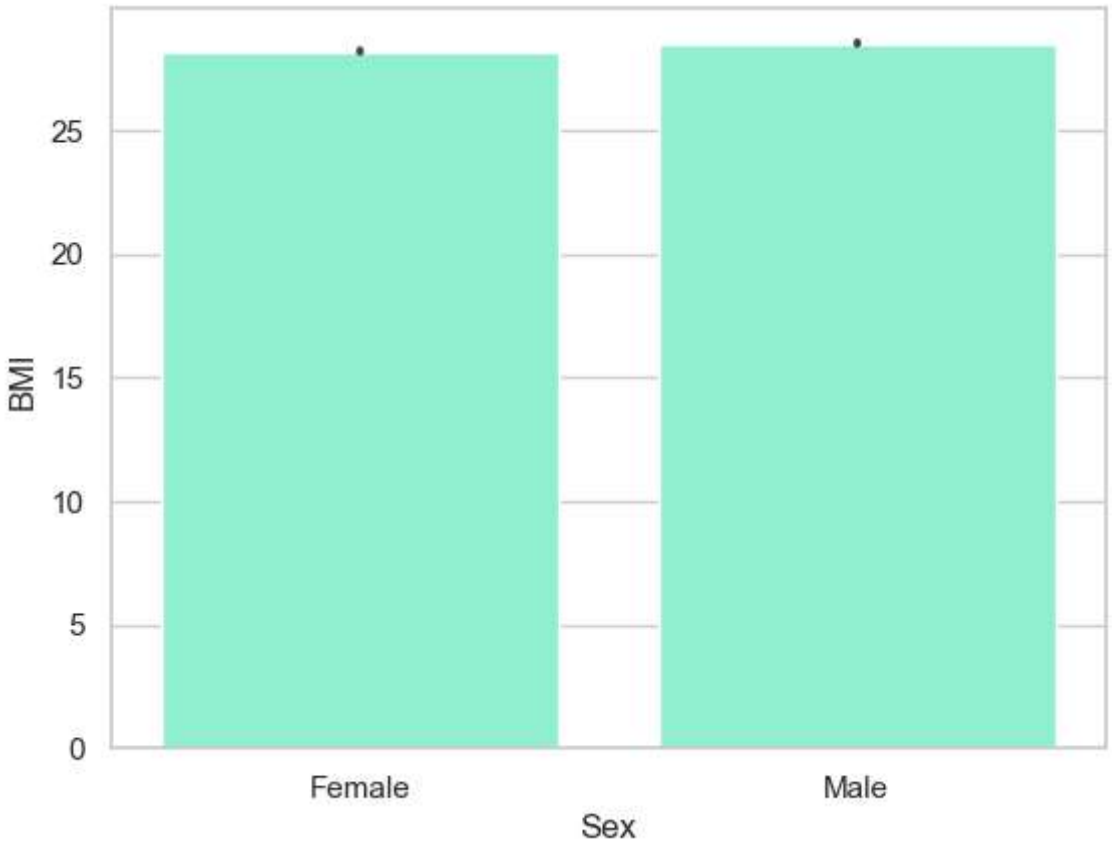
Out[7]:
```
HeartDisease        0
BMI                 0
Smoking             0
AlcoholDrinking     0
Stroke              0
PhysicalHealth      0
MentalHealth        0
DiffWalking         0
Sex                 0
AgeCategory         0
Race                0
Diabetic            0
PhysicalActivity    0
GenHealth           0
SleepTime           0
Asthma              0
KidneyDisease       0
SkinCancer          0
dtype: int64
```

In [8]:
```
df.head()
```

Out[8]:

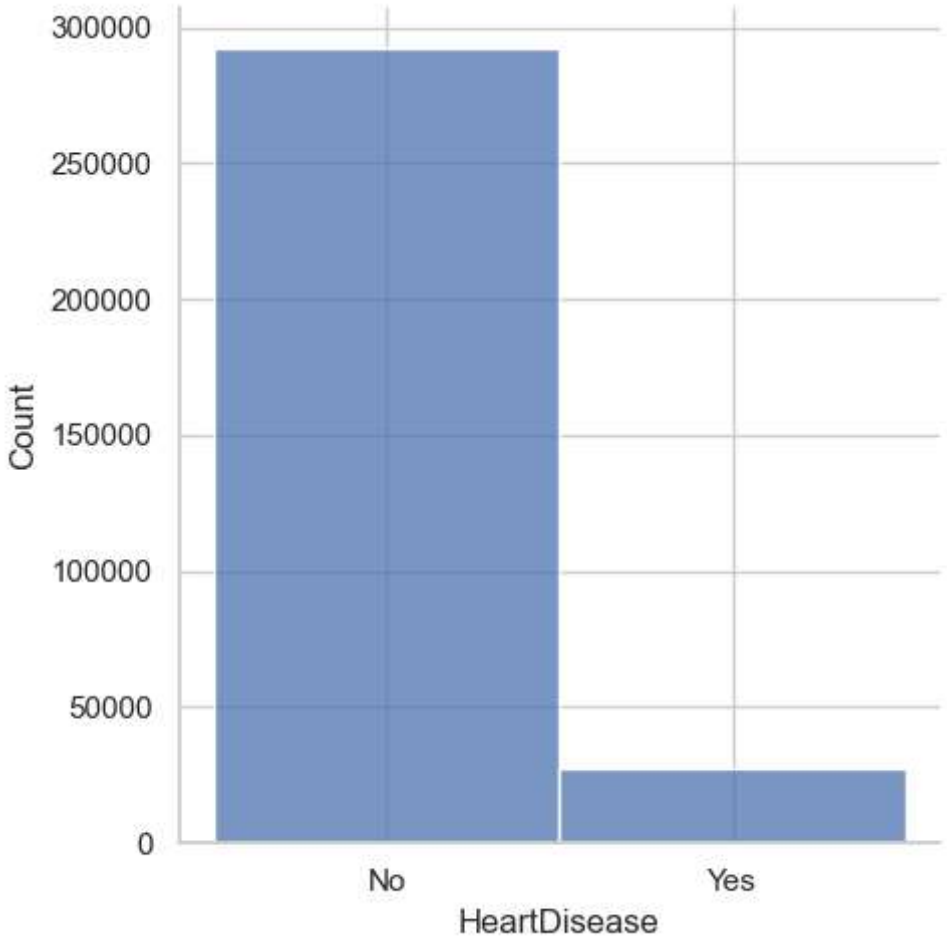
	HeartDisease	BMI	Smoking	AlcoholDrinking	Stroke	PhysicalHealth	MentalHealth	DiffWal
0	No	16.60	Yes	No	No	3.0	30.0	
1	No	20.34	No	No	Yes	0.0	0.0	
2	No	26.58	Yes	No	No	20.0	30.0	
3	No	24.21	No	No	No	0.0	0.0	
4	No	23.71	No	No	No	28.0	0.0	

In [9]:
```
sns.barplot(x='Sex',y='BMI',data=df,color='aquamarine')
plt.show()
```



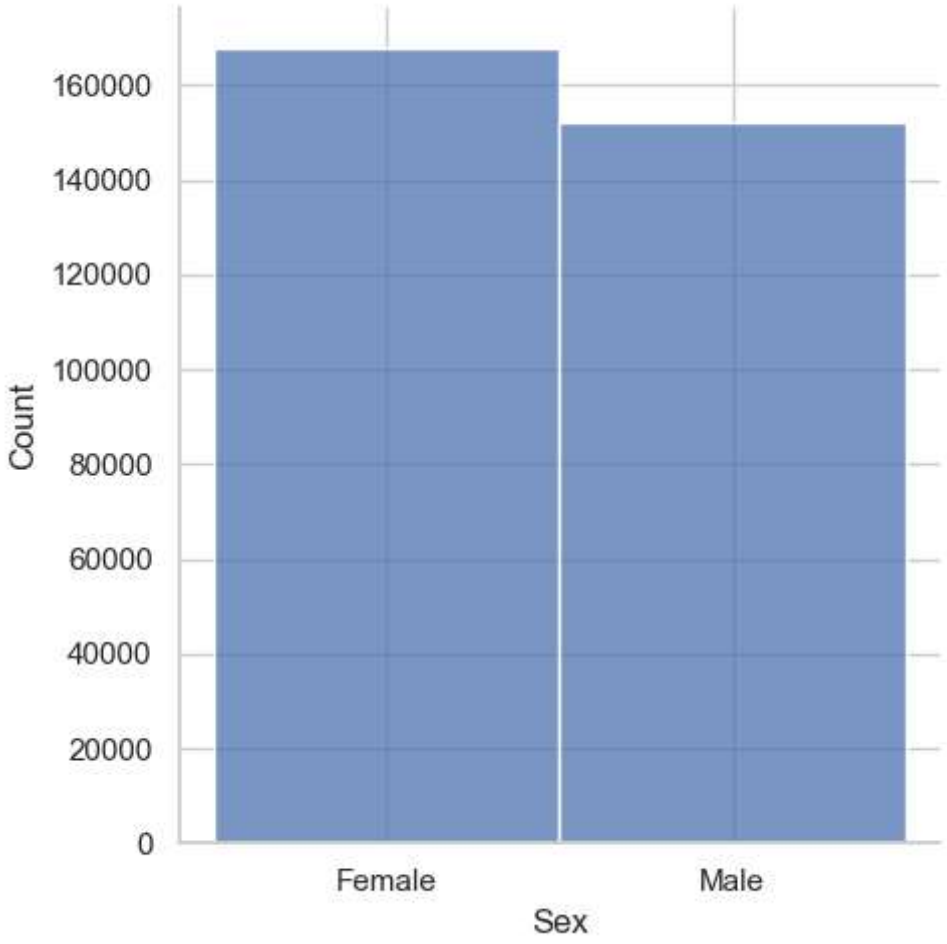
In [10]:
```
sns.displot(df['HeartDisease'])
```

Out[10]: <seaborn.axisgrid.FacetGrid at 0x1fa37fe2500>



In [11]:
```
sns.displot(df['Sex'])
```

Out[11]: <seaborn.axisgrid.FacetGrid at 0x1fa694de1a0>

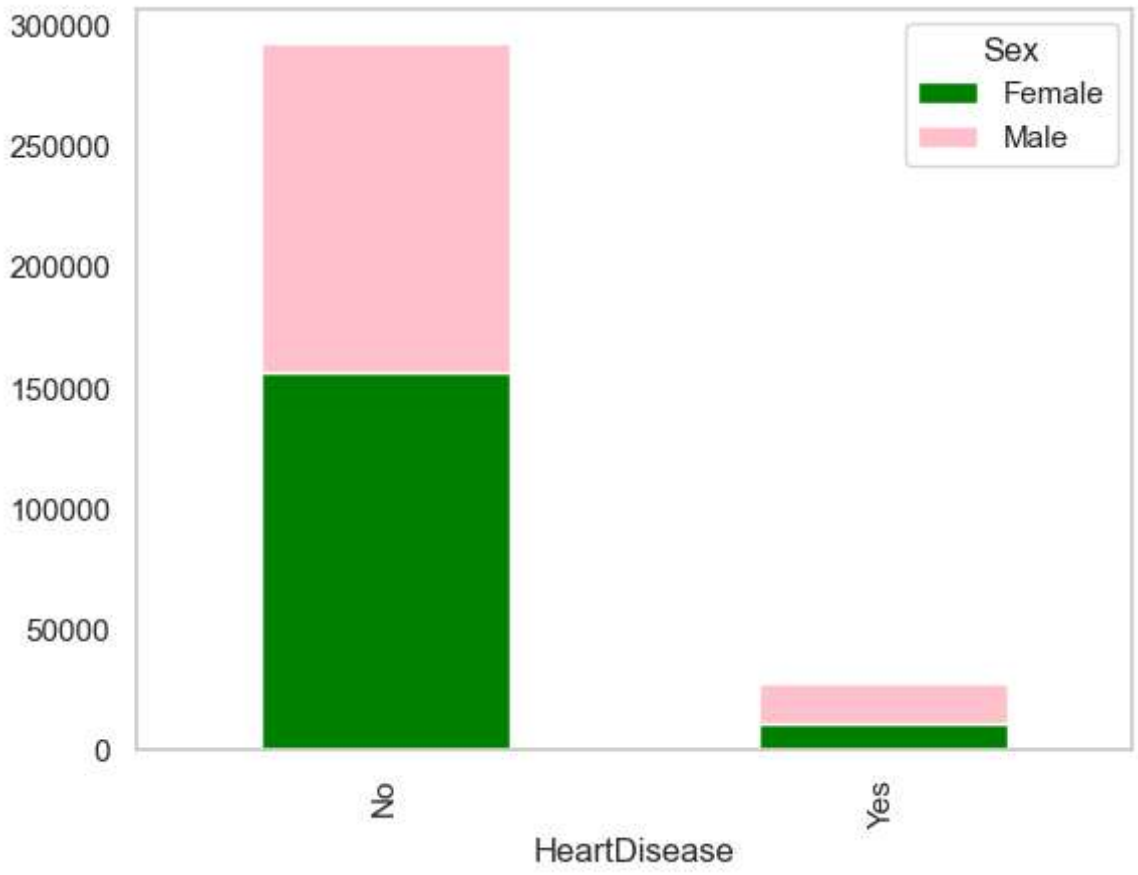


In [12]:
```
x=pd.crosstab(df['HeartDisease'],df['Sex'])
print(x)
```

```
Sex           Female    Male
HeartDisease
No            156571  135851
Yes            11234   16139
```

In [13]:
```
x.plot(kind='bar',stacked=True,color=['green','pink'],grid=False)
```

Out[13]: <Axes: xlabel='HeartDisease'>



conclusion

The males has high chances of getting Heart Diseases

In [ ]: