HAP_EDA

February 26, 2022

[2]:
```
from ipynb.fs.full.HAP_DataProcessing import *
import matplotlib.pyplot as plt
import seaborn as sns
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 595 entries, 0 to 927
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       595 non-null    float64
 1   sex       595 non-null    float64
 2   cp        595 non-null    float64
 3   trestbps  595 non-null    float64
 4   chol      595 non-null    float64
 5   fbs       595 non-null    float64
 6   restecg   595 non-null    float64
 7   thalach   595 non-null    float64
 8   exang     595 non-null    float64
 9   oldpeak   595 non-null    float64
 10  slope     595 non-null    float64
 11  ca        595 non-null    float64
 12  thal      595 non-null    float64
 13  num       595 non-null    int64
dtypes: float64(13), int64(1)
memory usage: 69.7 KB
```

```
None
```

[33]:
```
hap_df = pd.read_csv('C:\BU\portfolio\HeartAttackPrediction\dataset/
 ↪heart_attack.csv')
data=hap_df
hap_df_eda=hap_df
hap_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 595 entries, 0 to 594
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
```

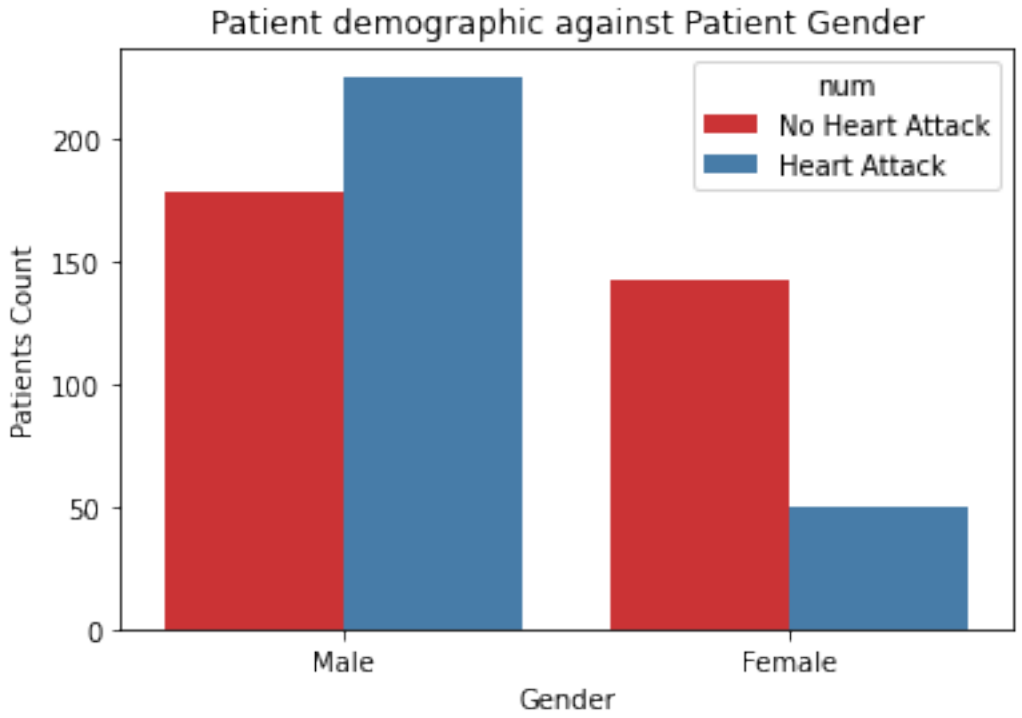
```
 0   age       595 non-null    float64
 1   sex       595 non-null    float64
 2   cp        595 non-null    float64
 3   trestbps  595 non-null    float64
 4   chol      595 non-null    float64
 5   fbs       595 non-null    float64
 6   restecg   595 non-null    float64
 7   thalach   595 non-null    float64
 8   exang     595 non-null    float64
 9   oldpeak   595 non-null    float64
 10  slope     595 non-null    float64
 11  ca        595 non-null    float64
 12  thal      595 non-null    float64
 13  num       595 non-null    int64
dtypes: float64(13), int64(1)
memory usage: 65.2 KB
```

[34]:
```
d= {1:"Heart Attack",0:"No Heart Attack"}
hap_df_eda.num = data.num.map(d)
s= {1:"Male",0:"Female"}
hap_df_eda.sex = data.sex.map(s)
c ={1:"typical angina",2:"atypical angina",3:"non-anginal pain",4:
 ↪"asymptomatic"}
hap_df_eda.cp = data.cp.map(c)
```

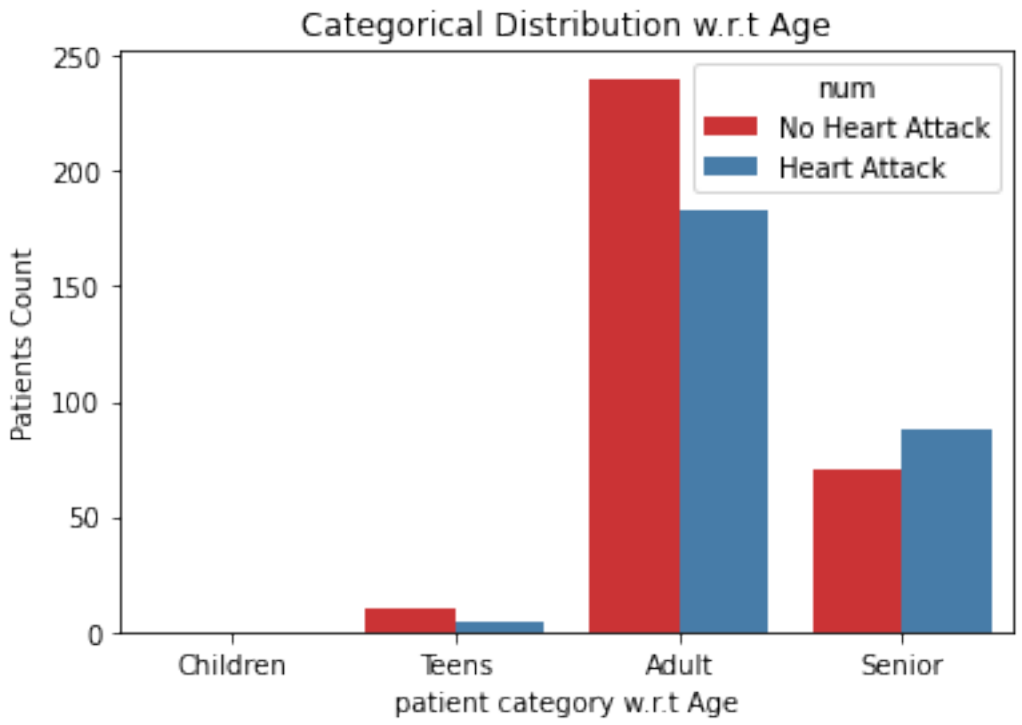
[35]:
```
hap_df_eda.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 595 entries, 0 to 594
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       595 non-null    float64
 1   sex       595 non-null    object
 2   cp        595 non-null    object
 3   trestbps  595 non-null    float64
 4   chol      595 non-null    float64
 5   fbs       595 non-null    float64
 6   restecg   595 non-null    float64
 7   thalach   595 non-null    float64
 8   exang     595 non-null    float64
 9   oldpeak   595 non-null    float64
 10  slope     595 non-null    float64
 11  ca        595 non-null    float64
 12  thal      595 non-null    float64
 13  num       595 non-null    object
dtypes: float64(11), object(3)
memory usage: 65.2+ KB
```

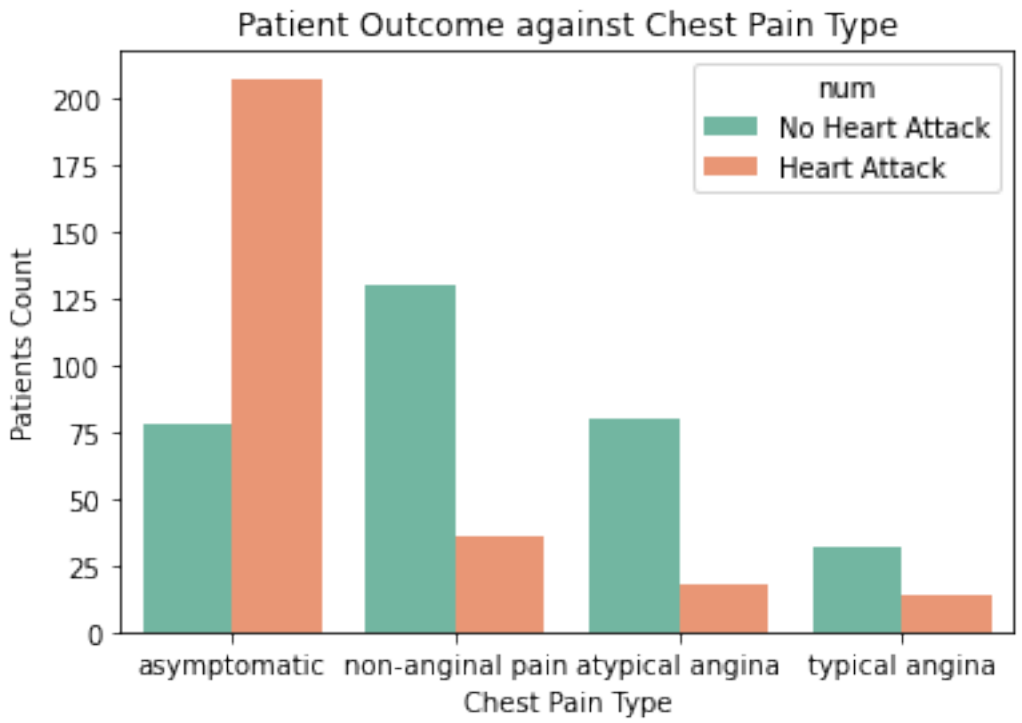
[36]:
```
ax = sns.countplot(x = 'sex', hue = 'num', palette = 'Set1', data = hap_df_eda)
ax.set(title = 'Patient demographic against Patient Gender',
       xlabel = 'Gender', ylabel = 'Patients Count')

plt.show()
```



[37]:
```
interval = (0,18,35,60,120)
categories = ['Children','Teens','Adult', 'Senior']
hap_df_eda['age'] = pd.cut(hap_df_eda.age, interval, labels = categories)

ax = sns.countplot(x = 'age',  data = hap_df_eda, hue = 'num', palette = 'Set1')

ax.set(xlabel='patient category w.r.t Age', ylabel='Patients Count',
       title="Categorical Distribution w.r.t Age")

plt.show()
```

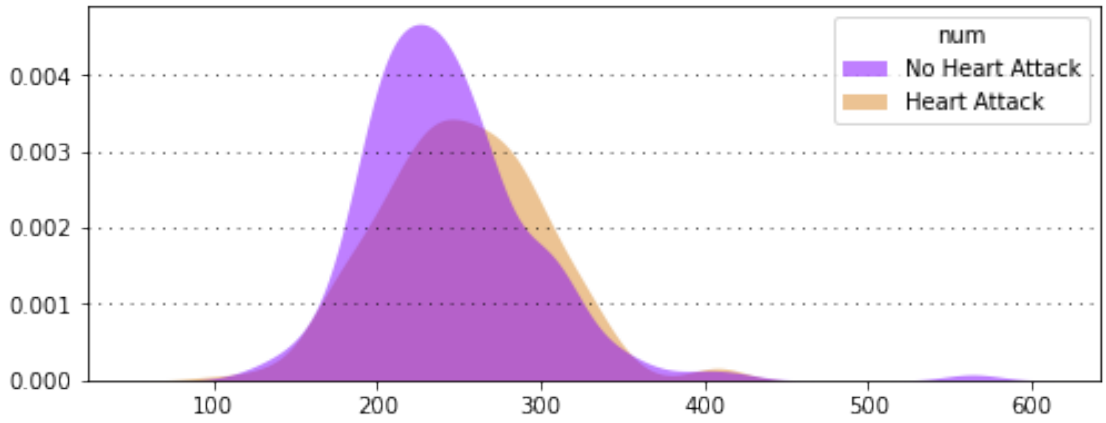


[45]:
```
ax = sns.countplot(x = 'cp', hue = 'num', palette = 'Set2', data = hap_df_eda)
ax.set(title = 'Patient Outcome against Chest Pain Type',
       xlabel = 'Chest Pain Type', ylabel = 'Patients Count')

plt.show()
```

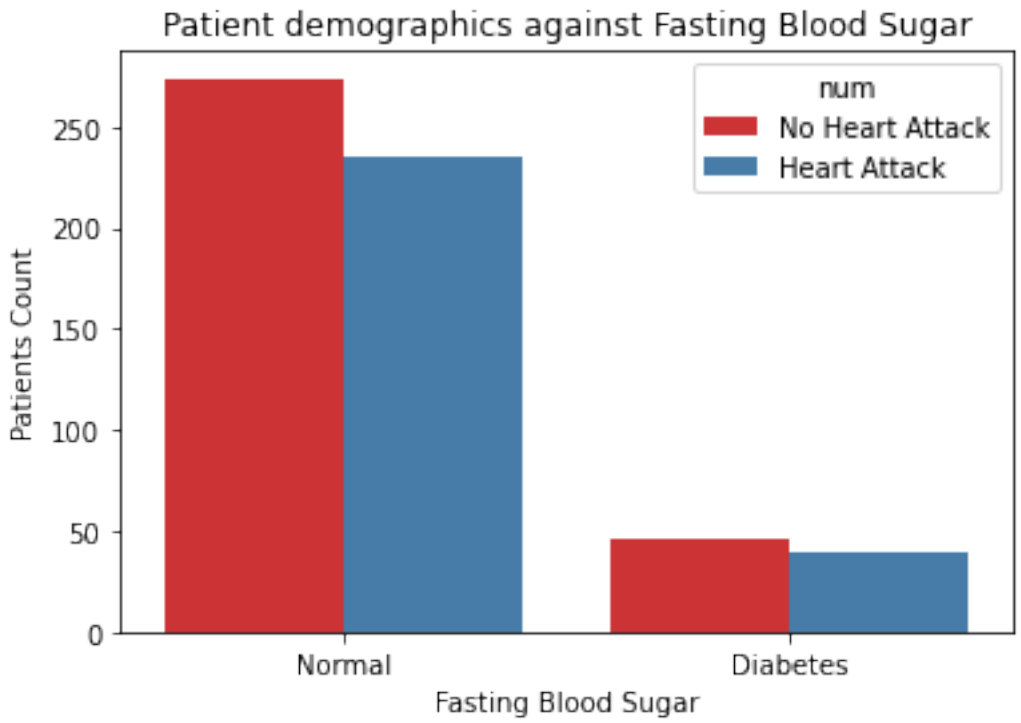


[42]:
```
fig = plt.figure(figsize=(18,18))
gs = fig.add_gridspec(5,2)
ax = fig.add_subplot(gs[2,1])

# Chol
ax.grid(color='#000000', linestyle=':', axis='y', zorder=0,  dashes=(1,5))
sns.kdeplot(ax=ax, data=hap_df_eda, x='chol',hue="num",␣
 ↪fill=True,palette=["#8000ff","#da8829"], alpha=.5, linewidth=0)
ax.set_xlabel("")
ax.set_ylabel("")
```

[42]:
```
Text(0, 0.5, '')
```

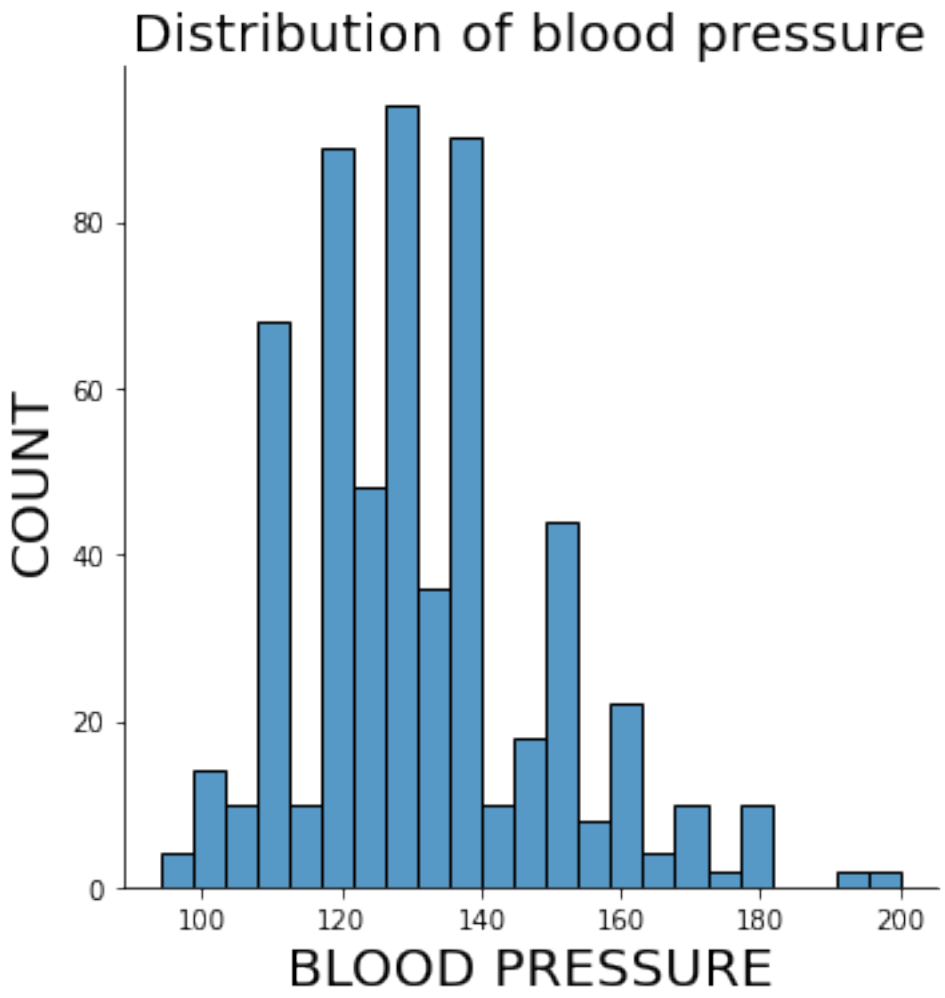


[43]:
```
s= {1:"Diabetes",0:"Normal"}
hap_df_eda.fbs = hap_df_eda.fbs.map(s)
ax = sns.countplot(x = 'fbs', hue = 'num', palette = 'Set1', data = hap_df_eda)
ax.set(title = 'Patient demographics against Fasting Blood Sugar',
       xlabel = 'Fasting Blood Sugar', ylabel = 'Patients Count')

plt.show()
```



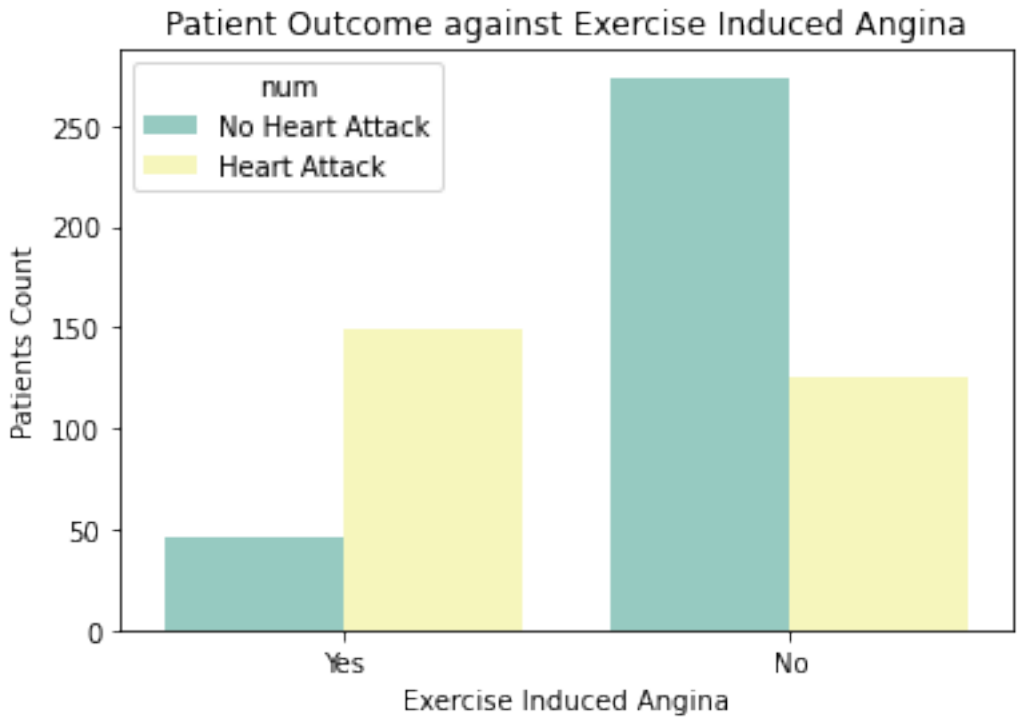
[44]:
```
plt.figure(figsize=(20,15))
sns.displot(hap_df_eda["trestbps"])
plt.title("Distribution of blood pressure",fontsize=20)
plt.xlabel("BLOOD PRESSURE",fontsize=20)
plt.ylabel("COUNT",fontsize=20)
plt.show()
```

```
<Figure size 1440x1080 with 0 Axes>
```

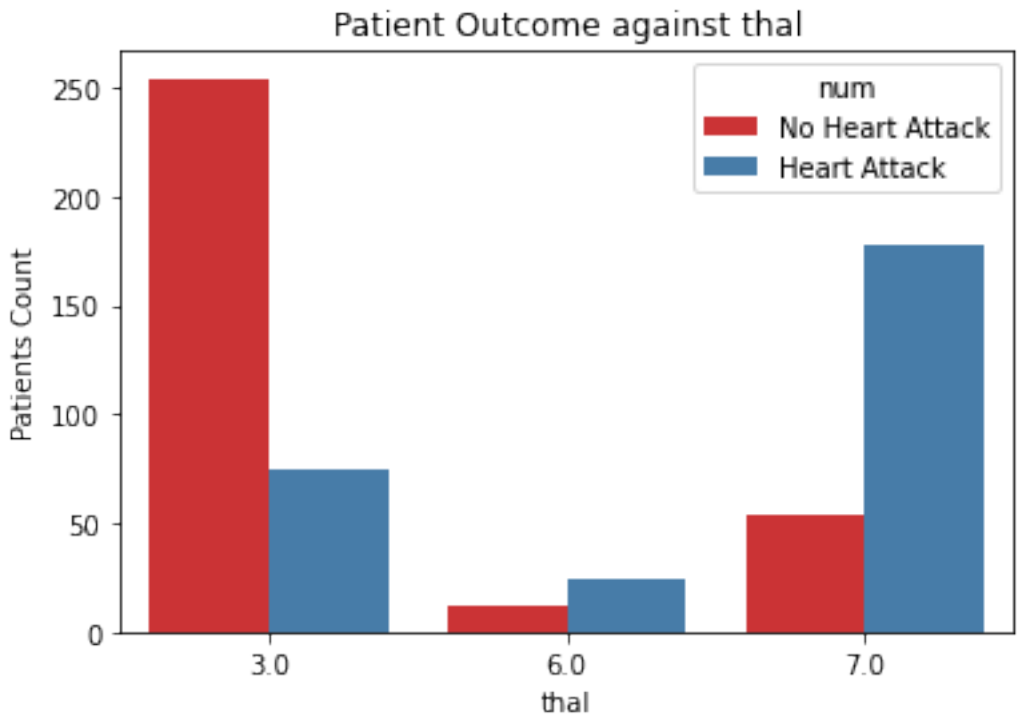


[46]:
```
s= {1:"Yes",0:"No"}
hap_df_eda.exang = hap_df_eda.exang.map(s)
ax = sns.countplot(x = 'exang', hue = 'num', palette = 'Set3', data =␣
 ↪hap_df_eda)
ax.set(title = 'Patient Outcome against Exercise Induced Angina',
       xlabel = 'Exercise Induced Angina', ylabel = 'Patients Count')
```

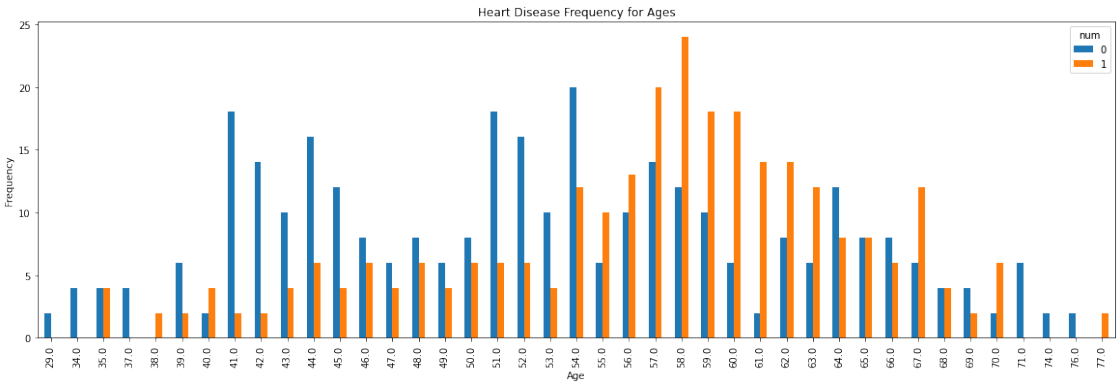
[46]:
```
[Text(0.5, 1.0, 'Patient Outcome against Exercise Induced Angina'),
 Text(0.5, 0, 'Exercise Induced Angina'),
 Text(0, 0.5, 'Patients Count')]
```



[48]:
```
ax = sns.countplot(x = 'thal', hue = 'num', palette = 'Set1', data = hap_df_eda)
ax.set(title = 'Patient Outcome against thal',
       xlabel = 'thal', ylabel = 'Patients Count')

plt.show()
```



[61]:
```
pd.crosstab(hap_df_eda.age,hap_df_eda.num).plot(kind="bar",figsize=(20,6))
plt.title('Heart Disease Frequency w.r.t Age')
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.show()
```



[67]:
```
plt.figure(figsize=(14,12))
sns.heatmap(hap_df.corr(),linewidths=0.1,
            square=True,  linecolor='white', annot=True)
```

```
plt.show()
```


age 1 -0.092 0.11 0.29 0.2 0.13 0.15 -0.39 0.2 0.16 0.36 0.13 0.23
sex -0.092 1 0.0099 -0.067 -0.2 0.038 0.033 -0.062 0.11 0.034 0.091 0.38 0.28
cp 0.11 0.0099 1 -0.038 0.067 -0.058 0.062 -0.34 0.2 0.15 0.23 0.27 0.41
trestbps 0.29 -0.067 -0.038 1 0.13 0.18 0.15 -0.048 0.19 0.12 0.099 0.14 0.15
chol 0.2 -0.2 0.067 0.13 1 0.015 0.17 0.006 0.037 -0.012 0.12 0.0052 0.075
fbs 0.13 0.038 -0.058 0.18 0.015 1 0.069 -0.0069 0.008 0.047 0.15 0.061 0.0024
restecg 0.15 0.033 0.062 0.15 0.17 0.069 1 -0.07 0.11 0.13 0.13 0.017 0.16
thalach -0.39 -0.062 -0.34 -0.048 0.006 -0.0069 -0.07 1 -0.35 -0.39 -0.27 -0.28 -0.43
oldpeak 0.2 0.11 0.2 0.19 0.037 0.008 0.11 -0.35 1 0.58 0.29 0.35 0.42
slope 0.16 0.034 0.15 0.12 -0.012 0.047 0.13 -0.39 0.58 1 0.11 0.28 0.33
ca 0.36 0.091 0.23 0.099 0.12 0.15 0.13 -0.27 0.29 0.11 1 0.25 0.46
thal 0.13 0.38 0.27 0.14 0.0052 0.061 0.017 -0.28 0.35 0.28 0.25 1 0.53
num 0.23 0.28 0.41 0.15 0.075 0.0024 0.16 -0.43 0.42 0.33 0.46 0.53 1
age sex cp trestbps chol fbs restecg thalach oldpeak slope ca thal num
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4

[ ]: