Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

jupyter Assign_1 (unsaved changes)

Logout

File Edit View Insert Cell Kernel Help | Not Trusted | Python 3 (ipykernel)

Run | Code

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

In [3]:
```
df = pd.read_csv("Heart.csv")
```

In [4]:
```
df.head(8)
```

Out[4]:

	Unnamed: 0	Age	Sex	ChestPain	RestBP	Chol	Fbs	RestECG	MaxHR	ExAng	Oldpeak	Slope	Ca	Thal	AHD
0	1	63	1	typical	145	233	1	2	150	0	2.3	3	0.0	fixed	No
1	2	67	1	asymptomatic	160	286	0	2	108	1	1.5	2	3.0	normal	Yes
2	3	67	1	asymptomatic	120	229	0	2	129	1	2.6	2	2.0	reversable	Yes
3	4	37	1	nonanginal	130	250	0	0	187	0	3.5	3	0.0	normal	No
4	5	41	0	nontypical	130	204	0	2	172	0	1.4	1	0.0	normal	No
5	6	56	1	nontypical	120	236	0	0	178	0	0.8	1	0.0	normal	No
6	7	62	0	asymptomatic	140	268	0	2	160	0	3.6	3	2.0	normal	Yes
7	8	57	0	asymptomatic	120	354	0	0	163	1	0.6	1	0.0	normal	No

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

Out[5]:
```
(303, 15)
```

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

Out[6]:
```
Unnamed: 0    0
Age           0
Sex           0
ChestPain     0
RestBP        0
Chol          0
Fbs           0
```

Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

jupyter Assign_1 (autosaved)

Logout

File Edit View Insert Cell Kernel Help | Not Trusted | Python 3 (ipykernel)

Run | Code

```
RestECG       0
MaxHR         0
ExAng         0
Oldpeak       0
Slope         0
Ca            4
Thal          2
AHD           0
dtype: int64
```

In [7]:
```
print("Total missing values: ", df.isnull().sum().sum())
```

```
Total missing values:  6
```

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 15 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Unnamed: 0  303 non-null    int64
 1   Age         303 non-null    int64
 2   Sex         303 non-null    int64
 3   ChestPain   303 non-null    object
 4   RestBP      303 non-null    int64
 5   Chol        303 non-null    int64
 6   Fbs         303 non-null    int64
 7   RestECG     303 non-null    int64
 8   MaxHR       303 non-null    int64
 9   ExAng       303 non-null    int64
 10  Oldpeak     303 non-null    float64
 11  Slope       303 non-null    int64
 12  Ca          299 non-null    float64
 13  Thal        301 non-null    object
 14  AHD         303 non-null    object
dtypes: float64(2), int64(10), object(3)
memory usage: 35.6+ KB
```

Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

jupyter Assign_1 (autosaved)

Logout

File Edit View Insert Cell Kernel Help | Not Trusted | Python 3 (ipykernel)

Run | Code

In [9]:
```
df.dtypes
```

Out[9]:
```
Unnamed: 0      int64
Age             int64
Sex             int64
ChestPain      object
RestBP          int64
Chol            int64
Fbs             int64
RestECG         int64
MaxHR           int64
ExAng           int64
Oldpeak       float64
Slope           int64
Ca            float64
Thal           object
AHD            object
dtype: object
```

In [10]:
```
(df == 0).sum(axis=0)
```

Out[10]:
```
Unnamed: 0      0
Age             0
Sex            97
ChestPain       0
RestBP          0
Chol            0
Fbs           258
RestECG       151
MaxHR           0
ExAng         204
Oldpeak        99
Slope           0
Ca            176
Thal            0
AHD             0
dtype: int64
```

Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

jupyter Assign_1 (autosaved)

Logout

File Edit View Insert Cell Kernel Help | Not Trusted | Python 3 (ipykernel)

Run | Code

In [11]:
```
mean_age =df['Age'].mean()
```

In [12]:
```
mean_age
```

Out[12]:
```
54.43894389438944
```

In [13]:
```
df.columns
```

Out[13]:
```
Index(['Unnamed: 0', 'Age', 'Sex', 'ChestPain', 'RestBP', 'Chol', 'Fbs',
       'RestECG', 'MaxHR', 'ExAng', 'Oldpeak', 'Slope', 'Ca', 'Thal', 'AHD'],
      dtype='object')
```

In [14]:
```
df2 = df.filter(['Age','Sex','ChestPain','RestBP','Chol'])
```

In [15]:
```
df2
```

Out[15]:

	Age	Sex	ChestPain	RestBP	Chol
0	63	1	typical	145	233
1	67	1	asymptomatic	160	286
2	67	1	asymptomatic	120	229
3	37	1	nonanginal	130	250
4	41	0	nontypical	130	204
...	...	...	...	...	...
298	45	1	typical	110	264
299	68	1	asymptomatic	144	193
300	57	1	asymptomatic	130	131
301	57	0	nontypical	130	236
302	38	1	nonanginal	138	175

303 rows × 5 columns

In [16]:
```
mean = df['Ca'].mean()
```

Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

jupyter Assign_1 (autosaved) Logout

File Edit View Insert Cell Kernel Help Not Trusted Python 3 (ipykernel)

Run Code

In [16]:
```
mean = df['Ca'].mean()
df['Ca'].fillna(value=mean,inplace=True)
```

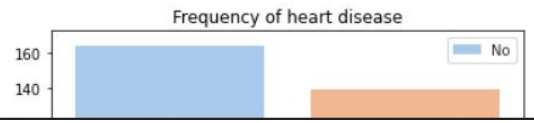
In [17]:
```
mode = df['Thal'].mode().iloc[0]
df['Thal'].fillna(value=mode, inplace=True)
```

In [18]:
```
df.isnull().sum()
```

Out[18]:
```
Unnamed: 0    0
Age           0
Sex           0
ChestPain     0
RestBP        0
Chol          0
Fbs           0
RestECG       0
MaxHR         0
ExAng         0
Oldpeak       0
Slope         0
Ca            0
Thal          0
AHD           0
dtype: int64
```

In [19]:
```
sns.countplot(x='AHD',data=df, palette='pastel')
sns.set_style("whitegrid")
plt.xlabel("Acute Heart Disease")
plt.ylabel("Total count")
plt.title("Frequency of heart disease")
plt.legend(['No','Yes'],loc='upper right')
```

Out[19]:
```
<matplotlib.legend.Legend at 0x7f348fc928b0>
```



Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

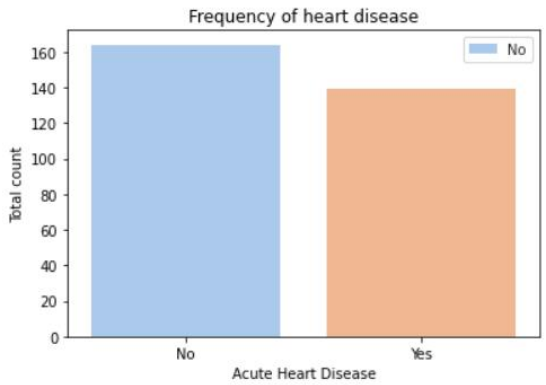
jupyter Assign_1 (autosaved) Logout

File Edit View Insert Cell Kernel Help Not Trusted Python 3 (ipykernel)

Run Code

In [19]:
```
sns.countplot(x='AHD',data=df, palette='pastel')
sns.set_style("whitegrid")
plt.xlabel("Acute Heart Disease")
plt.ylabel("Total count")
plt.title("Frequency of heart disease")
plt.legend(['No','Yes'],loc='upper right')
```

Out[19]: <matplotlib.legend.Legend at 0x7f348fc928b0>



In [20]:
```
fig, ax = plt.subplots()
fig.set_size_inches(10, 6)
sns.histplot(x="Age", data=df, hue="AHD", palette="bright")
sns.set_style("whitegrid")
plt.title("Effect of age on frequency of heart disease")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.legend(["Heart Disease","No Heart Disease"])
plt.show()
```

Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

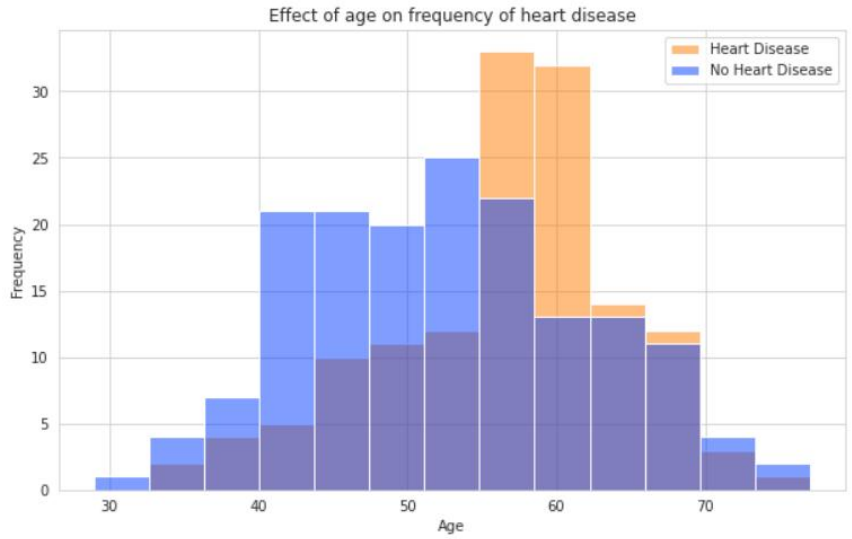
localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

jupyter Assign_1 (autosaved) Logout

File Edit View Insert Cell Kernel Help Not Trusted Python 3 (ipykernel)

Run Code

In [20]:
```
fig, ax = plt.subplots()
fig.set_size_inches(10, 6)
sns.histplot(x="Age", data=df, hue="AHD", palette="bright")
sns.set_style("whitegrid")
plt.title("Effect of age on frequency of heart disease")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.legend(["Heart Disease","No Heart Disease"])
plt.show()
```



In [21]:
```
df['ChestPain'].unique()
```

Out[21]:
```
array(['typical', 'asymptomatic', 'nonanginal', 'nontypical'],
      dtype=object)
```

Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

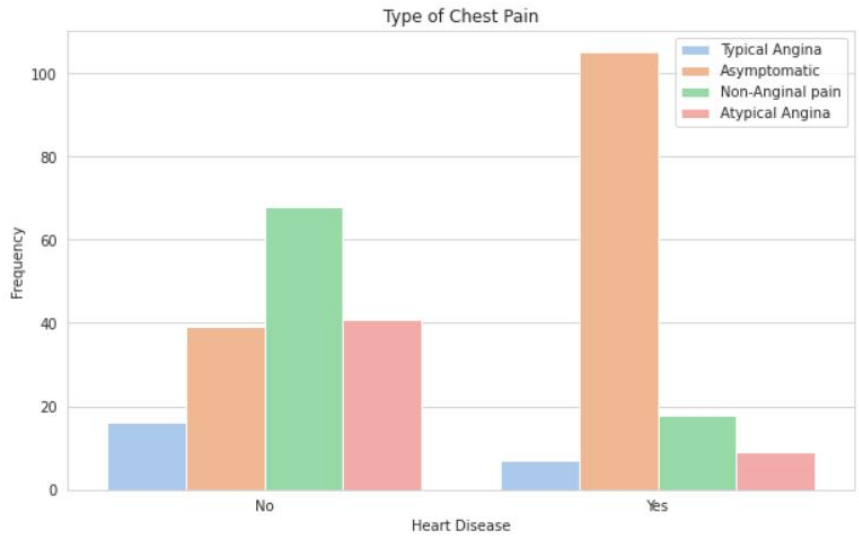
localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

jupyter Assign_1 (autosaved) Logout

File Edit View Insert Cell Kernel Help Not Trusted Python 3 (ipykernel)

Run Code

In [22]:
```
fig, ax = plt.subplots()
fig.set_size_inches(10, 6)
sns.countplot(x="AHD", hue="ChestPain", data=df, palette="pastel")
plt.title("Type of Chest Pain")
plt.xlabel("Heart Disease")
plt.ylabel("Frequency")
plt.legend(["Typical Angina","Asymptomatic", "Non-Anginal pain",
"Atypical Angina"])
plt.show()
```



In [23]:
```
fig, ax = plt.subplots()
fig.set_size_inches(10, 6)
sns.histplot(x="RestBP" , data=df, hue="AHD", palette="bright")
sns.set_style("whitegrid")
plt.title("Resting Blood Pressure")
```

Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

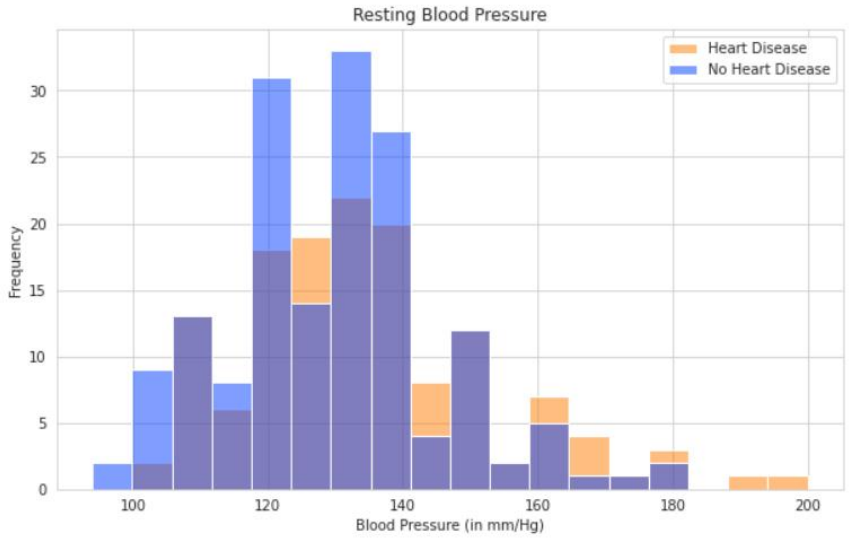
localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

jupyter Assign_1 (autosaved) Logout

File Edit View Insert Cell Kernel Help Not Trusted Python 3 (ipykernel)

Run Code

In [23]:
```
fig, ax = plt.subplots()
fig.set_size_inches(10, 6)
sns.histplot(x="RestBP" , data=df, hue="AHD", palette="bright")
sns.set_style("whitegrid")
plt.title("Resting Blood Pressure")
plt.xlabel("Blood Pressure (in mm/Hg)")
plt.ylabel("Frequency")
plt.legend(["Heart Disease","No Heart Disease"])
plt.show()
```



In [24]:
```
fig, ax = plt.subplots()
fig.set_size_inches(10, 6)
sns.histplot(x="Chol" , data=df, hue="AHD", palette="bright")
sns.set_style("whitegrid")
plt.title("Serum Cholesterol")
```

Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

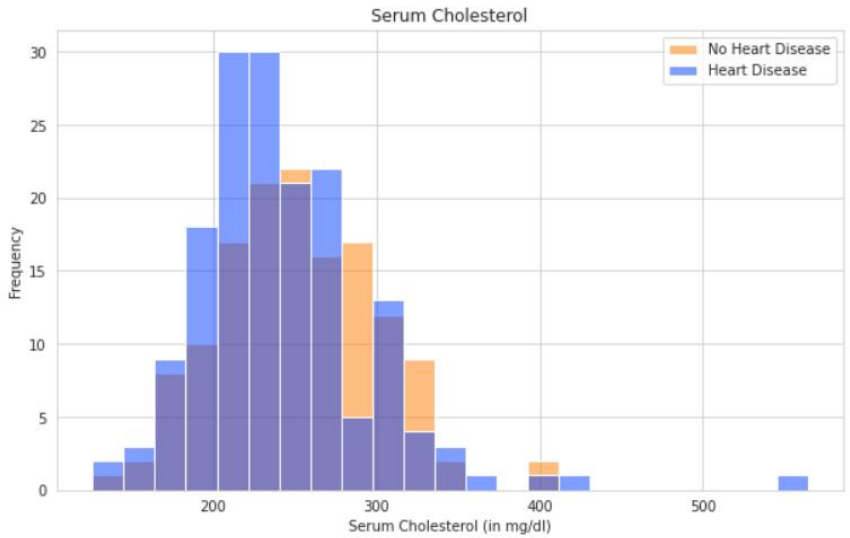
localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

jupyter Assign_1 (autosaved) Logout

File Edit View Insert Cell Kernel Help Not Trusted Python 3 (ipykernel)

Run Code

In [24]:
```
fig, ax = plt.subplots()
fig.set_size_inches(10, 6)
sns.histplot(x="Chol" , data=df, hue="AHD", palette="bright")
sns.set_style("whitegrid")
plt.title("Serum Cholesterol")
plt.xlabel("Serum Cholesterol (in mg/dl)")
plt.ylabel("Frequency")
plt.legend(["No Heart Disease","Heart Disease"])
plt.show()
```



In [26]:
```
fig, ax = plt.subplots()
fig.set_size_inches(10, 6)
sns.countplot(x="AHD", hue="RestECG", data=df, palette="pastel")
plt.title("Resting Electrocardiograph Results")
plt.xlabel("Heart Disease")
```

Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

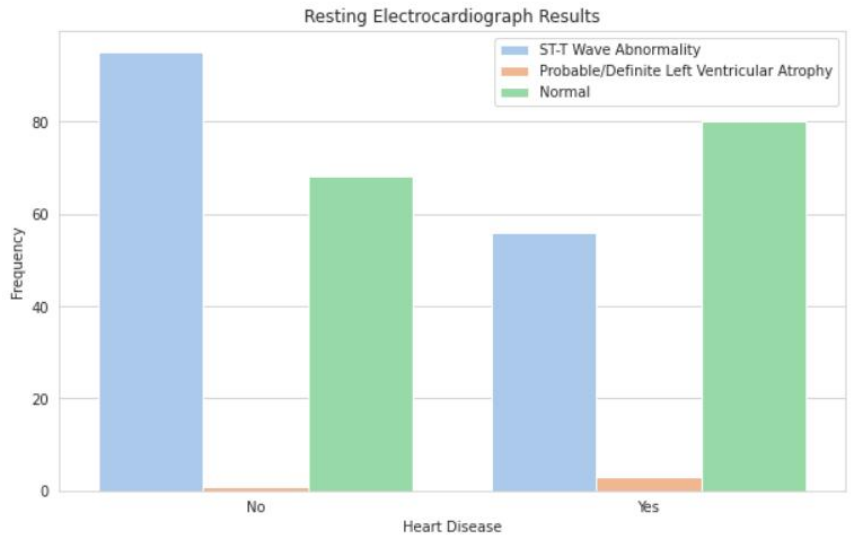
jupyter Assign_1 (autosaved) Logout

File Edit View Insert Cell Kernel Help Not Trusted Python 3 (ipykernel)

Run Code

Serum Cholesterol (in mg/dl)

In [26]:
```
fig, ax = plt.subplots()
fig.set_size_inches(10, 6)
sns.countplot(x="AHD", hue="RestECG", data=df, palette="pastel")
plt.title("Resting Electrocardiograph Results")
plt.xlabel("Heart Disease")
plt.ylabel("Frequency")
plt.legend(["ST-T Wave Abnormality", "Probable/Definite Left Ventricular Atrophy", "Normal"])
plt.show()
```



In [27]:
```
X = df[['Age','Sex','ChestPain','RestBP','Chol','RestECG','MaxHR']]
Y= df['AHD']
```

In [29]:
```
from sklearn.model_selection import train_test_split
```

Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

jupyter Assign_1 (autosaved) Logout

File Edit View Insert Cell Kernel Help Not Trusted Python 3 (ipykernel)

Run Code

In [27]:
```
X = df[['Age','Sex','ChestPain','RestBP','Chol','RestECG','MaxHR']]
Y= df['AHD']
```

In [29]:
```
from sklearn.model_selection import train_test_split
X_train, Y_train, X_test, Y_test = train_test_split(X,Y,test_size=0.25)
```

In [30]:
```
X_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 227 entries, 203 to 79
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Age        227 non-null    int64
 1   Sex        227 non-null    int64
 2   ChestPain  227 non-null    object
 3   RestBP     227 non-null    int64
 4   Chol       227 non-null    int64
 5   RestECG    227 non-null    int64
 6   MaxHR      227 non-null    int64
dtypes: int64(6), object(1)
memory usage: 14.2+ KB
```

In [31]:
```
Y_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 76 entries, 205 to 271
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   Age        76 non-null     int64
 1   Sex        76 non-null     int64
 2   ChestPain  76 non-null     object
 3   RestBP     76 non-null     int64
 4   Chol       76 non-null     int64
 5   RestECG    76 non-null     int64
 6   MaxHR      76 non-null     int64
dtypes: int64(6), object(1)
```

Desktop/ML/ML_LAB/ | Assign_1 - Jupyter Notebook

localhost:8888/notebooks/Desktop/ML/ML_LAB/Assign_1.ipynb

jupyter Assign_1 (autosaved) Logout

File Edit View Insert Cell Kernel Help Not Trusted Python 3 (ipykernel)

Run Code

In [30]:
```
X_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 227 entries, 203 to 79
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   Age        227 non-null    int64 
 1   Sex        227 non-null    int64 
 2   ChestPain  227 non-null    object
 3   RestBP     227 non-null    int64 
 4   Chol       227 non-null    int64 
 5   RestECG    227 non-null    int64 
 6   MaxHR      227 non-null    int64 
dtypes: int64(6), object(1)
memory usage: 14.2+ KB
```

In [31]:
```
Y_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 76 entries, 205 to 271
Data columns (total 7 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   Age        76 non-null     int64 
 1   Sex        76 non-null     int64 
 2   ChestPain  76 non-null     object
 3   RestBP     76 non-null     int64 
 4   Chol       76 non-null     int64 
 5   RestECG    76 non-null     int64 
 6   MaxHR      76 non-null     int64 
dtypes: int64(6), object(1)
memory usage: 4.8+ KB
```

In [ ]:

In [ ]: