In [7]:
```
import pandas as pd
import numpy as np
df=pd.read_csv('clevelanda.csv')
df
```

Out[7]:

	age	gender	cp	trestbps	chol	fps	restecg	thalach	exang	oldpeak	slope	ca	thal
0	63	1	1	145	233	1	2	150	0	2.3	3	0	6
1	67	1	4	160	286	0	2	108	1	1.5	2	3	3
2	67	1	4	120	229	0	2	129	1	2.6	2	2	7
3	37	1	3	130	250	0	0	187	0	3.5	3	0	3
4	41	0	2	130	204	0	2	172	0	1.4	1	0	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...
298	45	1	1	110	264	0	0	132	0	1.2	2	0	7
299	68	1	4	144	193	1	0	141	0	3.4	2	2	7
300	57	1	4	130	131	0	0	115	1	1.2	2	1	7
301	57	0	2	130	236	0	2	174	0	0.0	2	1	3
302	38	1	3	138	175	0	0	173	0	0.0	1	?	3

303 rows × 14 columns

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
age         0
gender      0
cp          0
trestbps    0
chol        0
fps         0
restecg     0
thalach     0
exang       0
oldpeak     0
slope       0
ca          0
thal        0
class       0
dtype: int64
```

In [10]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 303 entries, 0 to 302
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   age       303 non-null    int64  
 1   gender    303 non-null    int64  
 2   cp        303 non-null    int64  
 3   trestbps  303 non-null    int64  
 4   chol      303 non-null    int64  
 5   fps       303 non-null    int64  
 6   restecg   303 non-null    int64  
 7   thalach   303 non-null    int64  
 8   exang     303 non-null    int64  
 9   oldpeak   303 non-null    float64
 10  slope     303 non-null    int64  
 11  ca        303 non-null    object 
 12  thal      303 non-null    object 
 13  class     303 non-null    int64  
dtypes: float64(1), int64(11), object(2)
memory usage: 33.3+ KB
```

In [11]:
```
df.iloc[0:5,0:5]
```

Out[11]:

	age	gender	cp	trestbps	chol
0	63	1	1	145	233
1	67	1	4	160	286
2	67	1	4	120	229
3	37	1	3	130	250
4	41	0	2	130	204

In [12]:
```
df.ca.unique()
```

Out[12]:
```
array(['0', '3', '2', '1', '?'], dtype=object)
```

In [15]:
```
df.ca.value_counts()
```

Out[15]:
```
ca
0    176
1     65
2     38
3     20
?      4
Name: count, dtype: int64
```

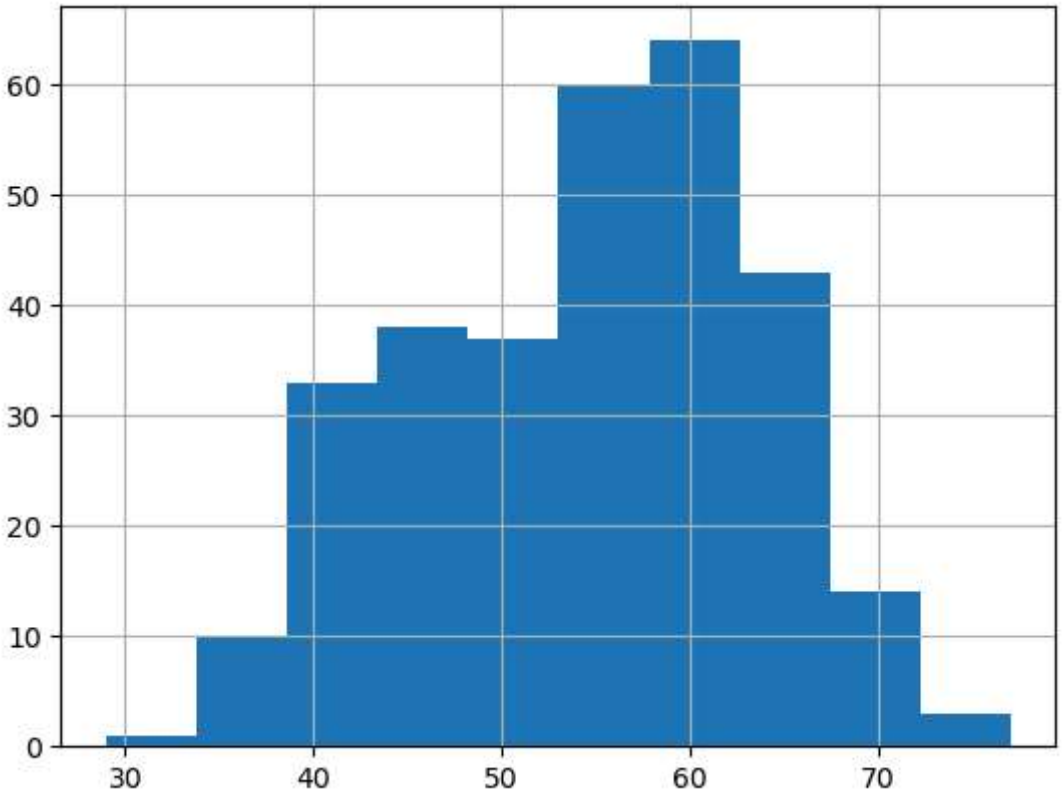
In [18]:
```
pip install plotly-express
```

```
Collecting plotly-express
  Obtaining dependency information for plotly-express from https://files.p
ythonhosted.org/packages/d4/d6/8a2906f51e073a4be80cab35cfa10e7a34853e60f3e
d5304ac470852a08d/plotly_express-0.4.1-py2.py3-none-any.whl.metadata (http
s://files.pythonhosted.org/packages/d4/d6/8a2906f51e073a4be80cab35cfa10e7a
34853e60f3ed5304ac470852a08d/plotly_express-0.4.1-py2.py3-none-any.whl.met
adata)
  Downloading plotly_express-0.4.1-py2.py3-none-any.whl.metadata (1.7 kB)
Requirement already satisfied: pandas>=0.20.0 in c:\users\cvr\anaconda3\li
b\site-packages (from plotly-express) (2.0.3)
Requirement already satisfied: plotly>=4.1.0 in c:\users\cvr\anaconda3\lib
\site-packages (from plotly-express) (5.9.0)
Requirement already satisfied: statsmodels>=0.9.0 in c:\users\cvr\anaconda
3\lib\site-packages (from plotly-express) (0.14.0)
Requirement already satisfied: scipy>=0.18 in c:\users\cvr\anaconda3\lib\s
ite-packages (from plotly-express) (1.11.1)
Requirement already satisfied: patsy>=0.5 in c:\users\cvr\anaconda3\lib\si
te-packages (from plotly-express) (0.5.3)
Requirement already satisfied: numpy>=1.11 in c:\users\cvr\anaconda3\lib\s
ite-packages (from plotly-express) (1.24.3)
Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\cvr\anac
onda3\lib\site-packages (from pandas>=0.20.0->plotly-express) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in c:\users\cvr\anaconda3\lib
\site-packages (from pandas>=0.20.0->plotly-express) (2023.3.post1)
Requirement already satisfied: tzdata>=2022.1 in c:\users\cvr\anaconda3\li
b\site-packages (from pandas>=0.20.0->plotly-express) (2023.3)
Requirement already satisfied: six in c:\users\cvr\anaconda3\lib\site-pack
ages (from patsy>=0.5->plotly-express) (1.16.0)
Requirement already satisfied: tenacity>=6.2.0 in c:\users\cvr\anaconda3\l
ib\site-packages (from plotly>=4.1.0->plotly-express) (8.2.2)
Requirement already satisfied: packaging>=21.3 in c:\users\cvr\anaconda3\l
ib\site-packages (from statsmodels>=0.9.0->plotly-express) (23.1)
Downloading plotly_express-0.4.1-py2.py3-none-any.whl (2.9 kB)
Installing collected packages: plotly-express
Successfully installed plotly-express-0.4.1
Note: you may need to restart the kernel to use updated packages.
```

In [26]:
```
import plotly.express as pt
p=pt.histogram(df,x='age',nbins=25)
p.show()
```

In [27]:
```
df['age'].hist()
```

Out[27]: <Axes: >



In [38]:
```
missing_locations = df[df.isnull().any(axis=1)]
print("\nRows with missing values:")
print(missing_locations)
filled_locations = missing_locations.fillna('NaN')
print("\nRows after filling missing values:")
print(filled_locations)
```

```
Rows with missing values:
Empty DataFrame
Columns: [age, gender, cp, trestbps, chol, fps, restecg, thalach, exang, o
ldpeak, slope, ca, thal, class]
Index: []

Rows after filling missing values:
Empty DataFrame
Columns: [age, gender, cp, trestbps, chol, fps, restecg, thalach, exang, o
ldpeak, slope, ca, thal, class]
Index: []
```

In [41]:
```
for i in df.columns:
    print(i,df[i].unique())
```

```
age [63 67 37 41 56 62 57 53 44 52 48 54 49 64 58 60 50 66 43 40 69 59 42
55
 61 65 71 51 46 45 39 68 47 34 35 29 70 77 38 74 76]
gender [1 0]
cp [1 4 3 2]
trestbps [145 160 120 130 140 172 150 110 132 117 135 112 105 124 125 142
128 170
 155 104 180 138 108 134 122 115 118 100 200  94 165 102 152 101 126 174
 148 178 158 192 129 144 123 136 146 106 156 154 114 164]
chol [233 286 229 250 204 236 268 354 254 203 192 294 256 263 199 168 239
275
 266 211 283 284 224 206 219 340 226 247 167 230 335 234 177 276 353 243
 225 302 212 330 175 417 197 198 290 253 172 273 213 305 216 304 188 282
 185 232 326 231 269 267 248 360 258 308 245 270 208 264 321 274 325 235
 257 164 141 252 255 201 222 260 182 303 265 309 307 249 186 341 183 407
 217 288 220 209 227 261 174 281 221 205 240 289 318 298 564 246 322 299
 300 293 277 214 207 223 160 394 184 315 409 244 195 196 126 313 259 200
 262 215 228 193 271 210 327 149 295 306 178 237 218 242 319 166 180 311
 278 342 169 187 157 176 241 131]
fps [1 0]
restecg [2 0 1]
thalach [150 108 129 187 172 178 160 163 147 155 148 153 142 173 162 174 1
68 139
 171 144 132 158 114 151 161 179 120 112 137 157 169 165 123 128 152 140
 188 109 125 131 170 113  99 177 141 180 111 143 182 156 115 149 145 146
 175 186 185 159 130 190 136  97 127 154 133 126 202 103 166 164 184 124
 122  96 138  88 105 194 195 106 167  95 192 117 121 116  71 118 181 134
  90]
exang [0 1]
oldpeak [2.3 1.5 2.6 3.5 1.4 0.8 3.6 0.6 3.1 0.4 1.3 0.  0.5 1.6 1.  1.2
0.2 1.8
 3.2 2.4 2.  2.5 2.2 2.8 3.  3.4 6.2 4.  5.6 2.9 0.1 2.1 1.9 4.2 0.9 1.1
 3.8 0.7 0.3 4.4]
slope [3 2 1]
ca ['0' '3' '2' '1' '?']
thal ['6' '3' '7' '?']
class [0 2 1 3 4]
```

In [39]:
```
df.ca.unique()
```

Out[39]:
```
array(['0', '3', '2', '1', '?'], dtype=object)
```

In [40]:
```
uni={col:df[col].unique() for col in df.columns}
uni
```

Out[40]:
```
{'age': array([63, 67, 37, 41, 56, 62, 57, 53, 44, 52, 48, 54, 49, 64, 58,
60, 50,
        66, 43, 40, 69, 59, 42, 55, 61, 65, 71, 51, 46, 45, 39, 68, 47, 3
4,
        35, 29, 70, 77, 38, 74, 76], dtype=int64),
 'gender': array([1, 0], dtype=int64),
 'cp': array([1, 4, 3, 2], dtype=int64),
 'trestbps': array([145, 160, 120, 130, 140, 172, 150, 110, 132, 117, 135,
112, 105,
        124, 125, 142, 128, 170, 155, 104, 180, 138, 108, 134, 122, 115,
        118, 100, 200,  94, 165, 102, 152, 101, 126, 174, 148, 178, 158,
        192, 129, 144, 123, 136, 146, 106, 156, 154, 114, 164], dtype=int6
4),
 'chol': array([233, 286, 229, 250, 204, 236, 268, 354, 254, 203, 192, 29
4, 256,
        263, 199, 168, 239, 275, 266, 211, 283, 284, 224, 206, 219, 340,
        226, 247, 167, 230, 335, 234, 177, 276, 353, 243, 225, 302, 212,
        330, 175, 417, 197, 198, 290, 253, 172, 273, 213, 305, 216, 304,
        188, 282, 185, 232, 326, 231, 269, 267, 248, 360, 258, 308, 245,
        270, 208, 264, 321, 274, 325, 235, 257, 164, 141, 252, 255, 201,
        222, 260, 182, 303, 265, 309, 307, 249, 186, 341, 183, 407, 217,
        288, 220, 209, 227, 261, 174, 281, 221, 205, 240, 289, 318, 298,
        564, 246, 322, 299, 300, 293, 277, 214, 207, 223, 160, 394, 184,
        315, 409, 244, 195, 196, 126, 313, 259, 200, 262, 215, 228, 193,
        271, 210, 327, 149, 295, 306, 178, 237, 218, 242, 319, 166, 180,
        311, 278, 342, 169, 187, 157, 176, 241, 131], dtype=int64),
 'fps': array([1, 0], dtype=int64),
 'restecg': array([2, 0, 1], dtype=int64),
 'thalach': array([150, 108, 129, 187, 172, 178, 160, 163, 147, 155, 148,
153, 142,
        173, 162, 174, 168, 139, 171, 144, 132, 158, 114, 151, 161, 179,
        120, 112, 137, 157, 169, 165, 123, 128, 152, 140, 188, 109, 125,
        131, 170, 113,  99, 177, 141, 180, 111, 143, 182, 156, 115, 149,
        145, 146, 175, 186, 185, 159, 130, 190, 136,  97, 127, 154, 133,
        126, 202, 103, 166, 164, 184, 124, 122,  96, 138,  88, 105, 194,
        195, 106, 167,  95, 192, 117, 121, 116,  71, 118, 181, 134,  90],
       dtype=int64),
 'exang': array([0, 1], dtype=int64),
 'oldpeak': array([2.3, 1.5, 2.6, 3.5, 1.4, 0.8, 3.6, 0.6, 3.1, 0.4, 1.3,
0. , 0.5,
        1.6, 1. , 1.2, 0.2, 1.8, 3.2, 2.4, 2. , 2.5, 2.2, 2.8, 3. , 3.4,
        6.2, 4. , 5.6, 2.9, 0.1, 2.1, 1.9, 4.2, 0.9, 1.1, 3.8, 0.7, 0.3,
        4.4]),
 'slope': array([3, 2, 1], dtype=int64),
 'ca': array(['0', '3', '2', '1', '?'], dtype=object),
 'thal': array(['6', '3', '7', '?'], dtype=object),
 'class': array([0, 2, 1, 3, 4], dtype=int64)}
```

In [42]:
```
df=df.replace('?',np.nan)
```

In [43]:
```
uni={col:df[col].unique() for col in df.columns}
uni
```

Out[43]:
```
{'age': array([63, 67, 37, 41, 56, 62, 57, 53, 44, 52, 48, 54, 49, 64, 58,
60, 50,
        66, 43, 40, 69, 59, 42, 55, 61, 65, 71, 51, 46, 45, 39, 68, 47, 3
4,
        35, 29, 70, 77, 38, 74, 76], dtype=int64),
 'gender': array([1, 0], dtype=int64),
 'cp': array([1, 4, 3, 2], dtype=int64),
 'trestbps': array([145, 160, 120, 130, 140, 172, 150, 110, 132, 117, 135,
112, 105,
        124, 125, 142, 128, 170, 155, 104, 180, 138, 108, 134, 122, 115,
        118, 100, 200,  94, 165, 102, 152, 101, 126, 174, 148, 178, 158,
        192, 129, 144, 123, 136, 146, 106, 156, 154, 114, 164], dtype=int6
4),
 'chol': array([233, 286, 229, 250, 204, 236, 268, 354, 254, 203, 192, 29
4, 256,
        263, 199, 168, 239, 275, 266, 211, 283, 284, 224, 206, 219, 340,
        226, 247, 167, 230, 335, 234, 177, 276, 353, 243, 225, 302, 212,
        330, 175, 417, 197, 198, 290, 253, 172, 273, 213, 305, 216, 304,
        188, 282, 185, 232, 326, 231, 269, 267, 248, 360, 258, 308, 245,
        270, 208, 264, 321, 274, 325, 235, 257, 164, 141, 252, 255, 201,
        222, 260, 182, 303, 265, 309, 307, 249, 186, 341, 183, 407, 217,
        288, 220, 209, 227, 261, 174, 281, 221, 205, 240, 289, 318, 298,
        564, 246, 322, 299, 300, 293, 277, 214, 207, 223, 160, 394, 184,
        315, 409, 244, 195, 196, 126, 313, 259, 200, 262, 215, 228, 193,
        271, 210, 327, 149, 295, 306, 178, 237, 218, 242, 319, 166, 180,
        311, 278, 342, 169, 187, 157, 176, 241, 131], dtype=int64),
 'fps': array([1, 0], dtype=int64),
 'restecg': array([2, 0, 1], dtype=int64),
 'thalach': array([150, 108, 129, 187, 172, 178, 160, 163, 147, 155, 148,
153, 142,
        173, 162, 174, 168, 139, 171, 144, 132, 158, 114, 151, 161, 179,
        120, 112, 137, 157, 169, 165, 123, 128, 152, 140, 188, 109, 125,
        131, 170, 113,  99, 177, 141, 180, 111, 143, 182, 156, 115, 149,
        145, 146, 175, 186, 185, 159, 130, 190, 136,  97, 127, 154, 133,
        126, 202, 103, 166, 164, 184, 124, 122,  96, 138,  88, 105, 194,
        195, 106, 167,  95, 192, 117, 121, 116,  71, 118, 181, 134,  90],
       dtype=int64),
 'exang': array([0, 1], dtype=int64),
 'oldpeak': array([2.3, 1.5, 2.6, 3.5, 1.4, 0.8, 3.6, 0.6, 3.1, 0.4, 1.3,
0. , 0.5,
        1.6, 1. , 1.2, 0.2, 1.8, 3.2, 2.4, 2. , 2.5, 2.2, 2.8, 3. , 3.4,
        6.2, 4. , 5.6, 2.9, 0.1, 2.1, 1.9, 4.2, 0.9, 1.1, 3.8, 0.7, 0.3,
        4.4]),
 'slope': array([3, 2, 1], dtype=int64),
 'ca': array(['0', '3', '2', '1', nan], dtype=object),
 'thal': array(['6', '3', '7', nan], dtype=object),
 'class': array([0, 2, 1, 3, 4], dtype=int64)}
```

In [46]:
```
df[['ca', 'thal']].mode()
```

	ca	thal
0	0	3

In [58]:
```
df['ca']=pd.to_numeric(df['ca'])
m=df['ca'].mean()
print(m)
df['thal']=pd.to_numeric(df['thal'])
n=df['thal'].mean()
print(n)
```

```
0.6722408026755853
4.73421926910299
```

In [66]:
```
mode_ca = df['ca'].mode()[0]  # Get the first mode value (if multiple modes

# Replace NaN values in the 'ca' column with the mode
df['ca'] = df['ca'].replace(np.nan, mode_ca)
mode_thal = df['thal'].mode()[0]  # Get the first mode value (if multiple mo

# Replace NaN values in the 'ca' column with the mode
df['thal'] = df['thal'].replace(np.nan, mode_thal)
```

In [67]:
```
uni={col:df[col].unique() for col in df.columns}
uni
```

Out[67]:
```
{'age': array([63, 67, 37, 41, 56, 62, 57, 53, 44, 52, 48, 54, 49, 64, 58,
60, 50,
        66, 43, 40, 69, 59, 42, 55, 61, 65, 71, 51, 46, 45, 39, 68, 47, 3
4,
        35, 29, 70, 77, 38, 74, 76], dtype=int64),
 'gender': array([1, 0], dtype=int64),
 'cp': array([1, 4, 3, 2], dtype=int64),
 'trestbps': array([145, 160, 120, 130, 140, 172, 150, 110, 132, 117, 135,
112, 105,
        124, 125, 142, 128, 170, 155, 104, 180, 138, 108, 134, 122, 115,
        118, 100, 200,  94, 165, 102, 152, 101, 126, 174, 148, 178, 158,
        192, 129, 144, 123, 136, 146, 106, 156, 154, 114, 164], dtype=int6
4),
 'chol': array([233, 286, 229, 250, 204, 236, 268, 354, 254, 203, 192, 29
4, 256,
        263, 199, 168, 239, 275, 266, 211, 283, 284, 224, 206, 219, 340,
        226, 247, 167, 230, 335, 234, 177, 276, 353, 243, 225, 302, 212,
        330, 175, 417, 197, 198, 290, 253, 172, 273, 213, 305, 216, 304,
        188, 282, 185, 232, 326, 231, 269, 267, 248, 360, 258, 308, 245,
        270, 208, 264, 321, 274, 325, 235, 257, 164, 141, 252, 255, 201,
        222, 260, 182, 303, 265, 309, 307, 249, 186, 341, 183, 407, 217,
        288, 220, 209, 227, 261, 174, 281, 221, 205, 240, 289, 318, 298,
        564, 246, 322, 299, 300, 293, 277, 214, 207, 223, 160, 394, 184,
        315, 409, 244, 195, 196, 126, 313, 259, 200, 262, 215, 228, 193,
        271, 210, 327, 149, 295, 306, 178, 237, 218, 242, 319, 166, 180,
        311, 278, 342, 169, 187, 157, 176, 241, 131], dtype=int64),
 'fps': array([1, 0], dtype=int64),
 'restecg': array([2, 0, 1], dtype=int64),
 'thalach': array([150, 108, 129, 187, 172, 178, 160, 163, 147, 155, 148,
153, 142,
        173, 162, 174, 168, 139, 171, 144, 132, 158, 114, 151, 161, 179,
        120, 112, 137, 157, 169, 165, 123, 128, 152, 140, 188, 109, 125,
        131, 170, 113,  99, 177, 141, 180, 111, 143, 182, 156, 115, 149,
        145, 146, 175, 186, 185, 159, 130, 190, 136,  97, 127, 154, 133,
        126, 202, 103, 166, 164, 184, 124, 122,  96, 138,  88, 105, 194,
        195, 106, 167,  95, 192, 117, 121, 116,  71, 118, 181, 134,  90],
       dtype=int64),
 'exang': array([0, 1], dtype=int64),
 'oldpeak': array([2.3, 1.5, 2.6, 3.5, 1.4, 0.8, 3.6, 0.6, 3.1, 0.4, 1.3,
0. , 0.5,
        1.6, 1. , 1.2, 0.2, 1.8, 3.2, 2.4, 2. , 2.5, 2.2, 2.8, 3. , 3.4,
        6.2, 4. , 5.6, 2.9, 0.1, 2.1, 1.9, 4.2, 0.9, 1.1, 3.8, 0.7, 0.3,
        4.4]),
 'slope': array([3, 2, 1], dtype=int64),
 'ca': array([0., 3., 2., 1.]),
 'thal': array([6., 3., 7.]),
 'class': array([0, 2, 1, 3, 4], dtype=int64)}
```

In [ ]:
```

```