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

In [2]:
```
df=pd.read_csv('heart.csv')
```

In [3]:
```
df
```

Out[3]:

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1
5	57	1	0	140	192	0	1	148	0	0.4	1	0	1	1
6	56	0	1	140	294	0	0	153	0	1.3	1	0	2	1
7	44	1	1	120	263	0	1	173	0	0.0	2	0	3	1
8	52	1	2	172	199	1	1	162	0	0.5	2	0	3	1
9	57	1	2	150	168	0	1	174	0	1.6	2	0	2	1
10	54	1	0	140	239	0	1	160	0	1.2	2	0	2	1
11	48	0	2	130	275	0	1	139	0	0.2	2	0	2	1
12	49	1	1	130	266	0	1	171	0	0.6	2	0	2	1
13	64	1	3	110	211	0	0	144	1	1.8	1	0	2	1
14	58	0	3	150	283	1	0	162	0	1.0	2	0	2	1

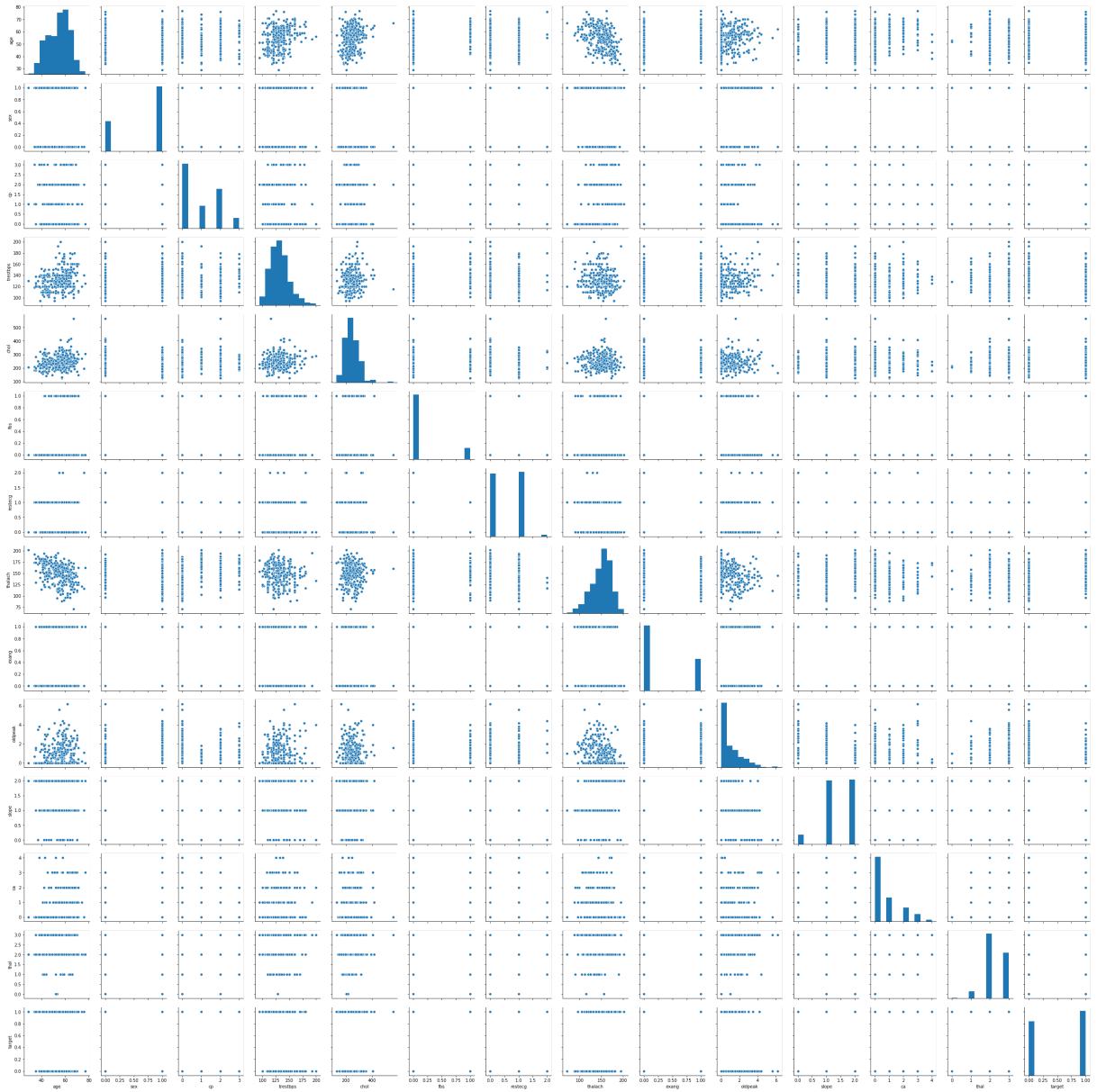
	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
15	50	0	2	120	219	0	1	158	0	1.6	1	0	2	1
16	58	0	2	120	340	0	1	172	0	0.0	2	0	2	1
17	66	0	3	150	226	0	1	114	0	2.6	0	0	2	1
18	43	1	0	150	247	0	1	171	0	1.5	2	0	2	1
19	69	0	3	140	239	0	1	151	0	1.8	2	2	2	1
20	59	1	0	135	234	0	1	161	0	0.5	1	0	3	1
21	44	1	2	130	233	0	1	179	1	0.4	2	0	2	1
22	42	1	0	140	226	0	1	178	0	0.0	2	0	2	1
23	61	1	2	150	243	1	1	137	1	1.0	1	0	2	1
24	40	1	3	140	199	0	1	178	1	1.4	2	0	3	1
25	71	0	1	160	302	0	1	162	0	0.4	2	2	2	1
26	59	1	2	150	212	1	1	157	0	1.6	2	0	2	1
27	51	1	2	110	175	0	1	123	0	0.6	2	0	2	1
28	65	0	2	140	417	1	0	157	0	0.8	2	1	2	1
29	53	1	2	130	197	1	0	152	0	1.2	0	0	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	..
273	58	1	0	100	234	0	1	156	0	0.1	2	1	3	0
274	47	1	0	110	275	0	0	118	1	1.0	1	1	2	0
275	52	1	0	125	212	0	1	168	0	1.0	2	2	3	0
276	58	1	0	146	218	0	1	105	0	2.0	1	1	3	0
277	57	1	1	124	261	0	1	141	0	0.3	2	0	3	0
278	58	0	1	136	319	1	0	152	0	0.0	2	2	2	0
279	61	1	0	138	166	0	0	125	1	3.6	1	1	2	0
280	42	1	0	136	315	0	1	125	1	1.8	1	0	1	0
281	52	1	0	128	204	1	1	156	1	1.0	1	0	0	0

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
282	59	1	2	126	218	1	1	134	0	2.2	1	1	1	0
283	40	1	0	152	223	0	1	181	0	0.0	2	0	3	0
284	61	1	0	140	207	0	0	138	1	1.9	2	1	3	0
285	46	1	0	140	311	0	1	120	1	1.8	1	2	3	0
286	59	1	3	134	204	0	1	162	0	0.8	2	2	2	0
287	57	1	1	154	232	0	0	164	0	0.0	2	1	2	0
288	57	1	0	110	335	0	1	143	1	3.0	1	1	3	0
289	55	0	0	128	205	0	2	130	1	2.0	1	1	3	0
290	61	1	0	148	203	0	1	161	0	0.0	2	1	3	0
291	58	1	0	114	318	0	2	140	0	4.4	0	3	1	0
292	58	0	0	170	225	1	0	146	1	2.8	1	2	1	0
293	67	1	2	152	212	0	0	150	0	0.8	1	0	3	0
294	44	1	0	120	169	0	1	144	1	2.8	0	0	1	0
295	63	1	0	140	187	0	0	144	1	4.0	2	2	3	0
296	63	0	0	124	197	0	1	136	1	0.0	1	0	2	0
297	59	1	0	164	176	1	0	90	0	1.0	1	2	1	0
298	57	0	0	140	241	0	1	123	1	0.2	1	0	3	0
299	45	1	3	110	264	0	1	132	0	1.2	1	0	3	0
300	68	1	0	144	193	1	1	141	0	3.4	1	2	3	0
301	57	1	0	130	131	0	1	115	1	1.2	1	1	3	0
302	57	0	1	130	236	0	0	174	0	0.0	1	1	2	0

303 rows × 14 columns

In [4]:
```
sns.pairplot(data=df)
```

Out[4]: <seaborn.axisgrid.PairGrid at 0x1f1f1ab8b00>



In [23]:
```
youth=df[(df['age']<=40) & (df['target']==1)]
```

```
len(youth)

middleage=df[(df.age>40) & (df.target==1)]
len(middleage)
```

Out[23]:
```
152
```

In [20]:
```
df.target.value_counts()
```

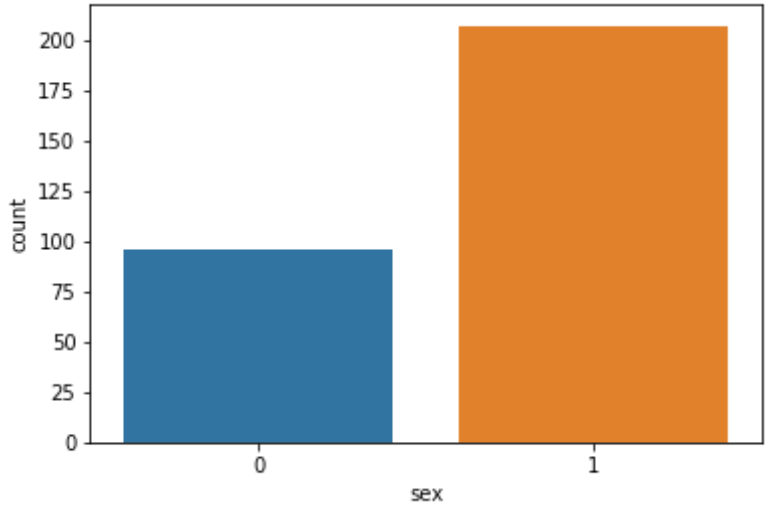
Out[20]:
```
1    165
0    138
Name: target, dtype: int64
```

In [21]:
```
df.sex.value_counts()
```

Out[21]:
```
1    207
0     96
Name: sex, dtype: int64
```

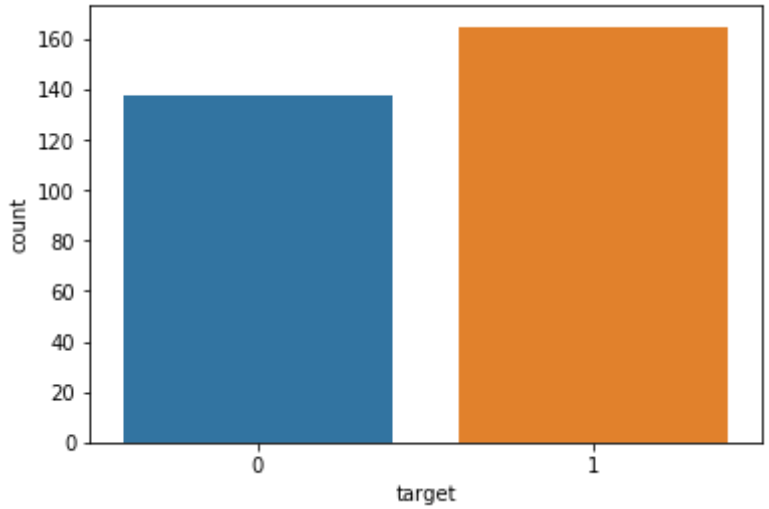
In [24]:
```
sns.countplot('sex',data=df)
```

Out[24]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1f1fd7dbdd8>
```



In [26]:
```
sns.countplot('target',data=df)
```

Out[26]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1f1fd7e9748>
```



In [27]:
```
df.columns
```

Out[27]:
```
Index(['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thala
ch',
       'exang', 'oldpeak', 'slope', 'ca', 'thal', 'target'],
      dtype='object')
```

In [28]:
```
X=df[['age', 'sex', 'cp', 'trestbps', 'chol', 'fbs', 'restecg', 'thalac
h',
       'exang', 'oldpeak', 'slope', 'ca', 'thal']]
```

In [29]:
```
y=df['target']
```

In [30]:
```
from sklearn.cross_validation import train_test_split
```

In [32]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4
, random_state=101)
```

In [33]:
```
from sklearn.linear_model import LogisticRegression
```

In [34]:
```
lm=LogisticRegression()
```

In [40]:
```
lm.fit(X_train,y_train)
```

Out[40]:
```
LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=
True,
          intercept_scaling=1, max_iter=100, multi_class='ovr', n_jobs=
1,
          penalty='l2', random_state=None, solver='liblinear', tol=0.00
01,
          verbose=0, warm_start=False)
```

In [ ]:
```

```

In [37]:
```
from sklearn.metrics import classification_report
```

In [39]:
```
from sklearn.metrics import confusion_matrix
```

In [42]:
```
predictions=lm.predict(X_test)
```

In [46]:
```
print(classification_report(y_test,predictions))
```

```
             precision    recall  f1-score   support

          0       0.92      0.77      0.84        60
          1       0.81      0.94      0.87        62

avg / total       0.86      0.85      0.85       122
```

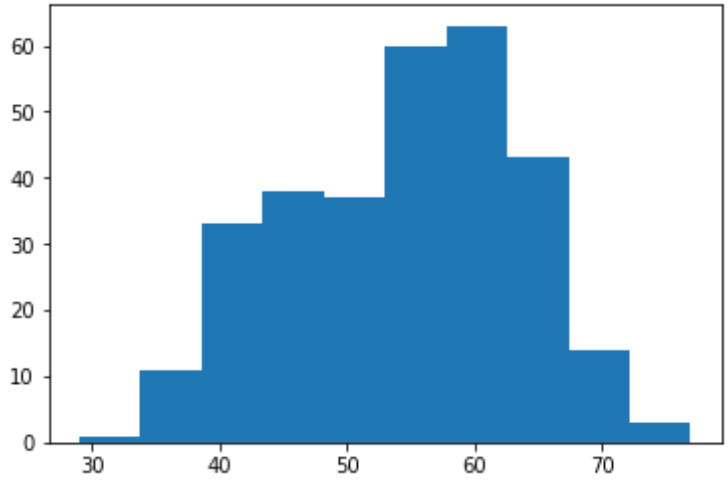
In [47]:
```
print(confusion_matrix(y_test,predictions))
```

```
[[46 14]
```

```
 [ 4 58]]
```

In [49]:
```
df.age.hist(grid=False)
```

Out[49]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1f1fdd623c8>
```



In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]:

In [ ]: