In [5]:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

In [6]:

```
autodv = pd.read_csv('x_automobile.csv')
```

In [7]:

```
autodv
```

Out[7]:

	symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location
0	3	122.0	alfa-romero	gas	std	2	convertible	rwd	front
1	3	122.0	alfa-romero	gas	std	2	convertible	rwd	front
2	1	122.0	alfa-romero	gas	std	2	hatchback	rwd	front
3	2	164.0	audi	gas	std	4	sedan	fwd	front
4	2	164.0	audi	gas	std	4	sedan	4wd	front
...	...	...	...	...	...	...	...	...	...
196	-1	95.0	volvo	gas	std	4	sedan	rwd	front
197	-1	95.0	volvo	gas	turbo	4	sedan	rwd	front
198	-1	95.0	volvo	gas	std	4	sedan	rwd	front
199	-1	95.0	volvo	diesel	turbo	4	sedan	rwd	front
200	-1	95.0	volvo	gas	turbo	4	sedan	rwd	front

201 rows × 26 columns

In [8]:

```
autodv.columns
```

Out[8]:

```
Index(['symboling', 'normalized-losses', 'make', 'fuel-type', 'aspiratio
n',
       'num-of-doors', 'body-style', 'drive-wheels', 'engine-location',
       'wheel-base', 'length', 'width', 'height', 'curb-weight', 'engine-t
ype',
       'num-of-cylinders', 'engine-size', 'fuel-system', 'bore', 'stroke',
       'compression-ratio', 'horsepower', 'peak-rpm', 'city-mpg',
       'highway-mpg', 'price'],
      dtype='object')
```

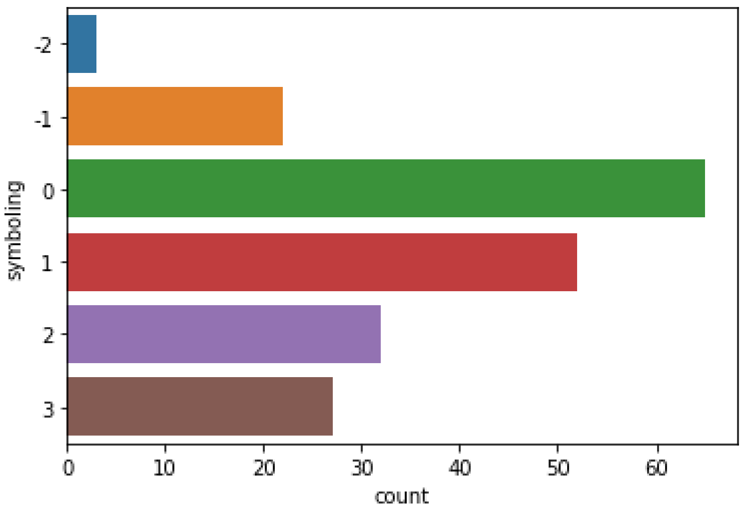
In [9]:

```
## Count plot

sns.countplot(y='symboling',data=autodv)
```

Out[9]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xa14fc48>
```



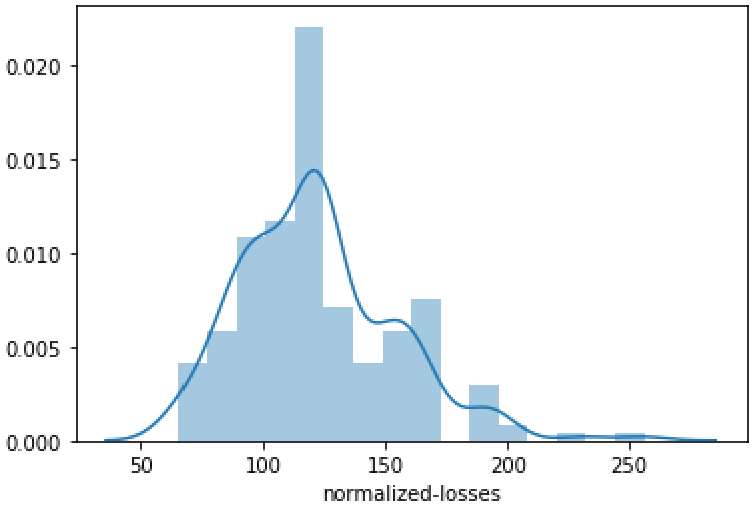
- +3 indicates high risk factor and -3 indicates safe
- Graph indicates no risk value(-3)
- safe count around 20-30

In [10]:

```
#dist plot helps to see colum feature distribution
sns.distplot(autodv['normalized-losses'])
```

Out[10]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xb8c9048>
```



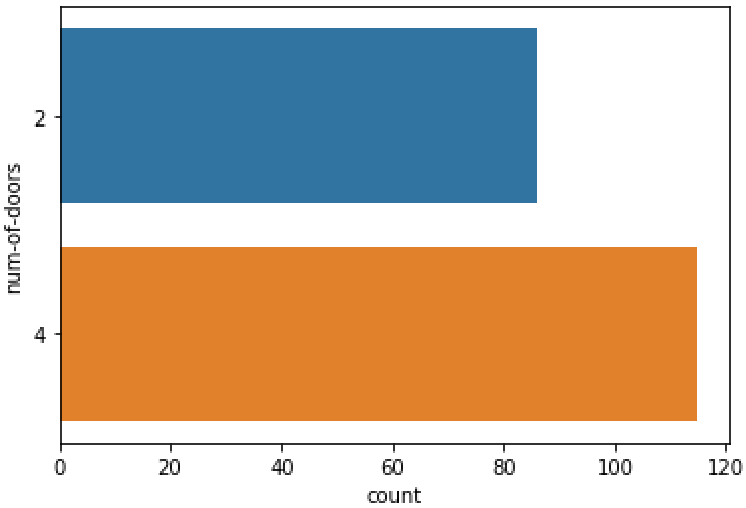
- normalized losses range around 60 to 180(assumption from graph)
- Graph shows high range from 100 -150

In [12]:

```
sns.countplot(y='num-of-doors',data=autodv)
```

Out[12]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xbb6f7c8>
```



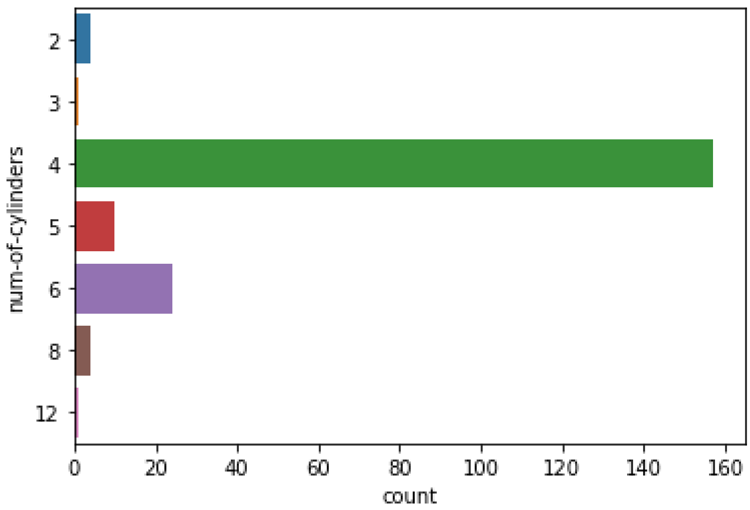
- cars with 4 doors are more compared to 2 doors from the graph

In [13]:

```
sns.countplot(y='num-of-cylinders',data=autodv)
```

Out[13]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xbbc0248>
```



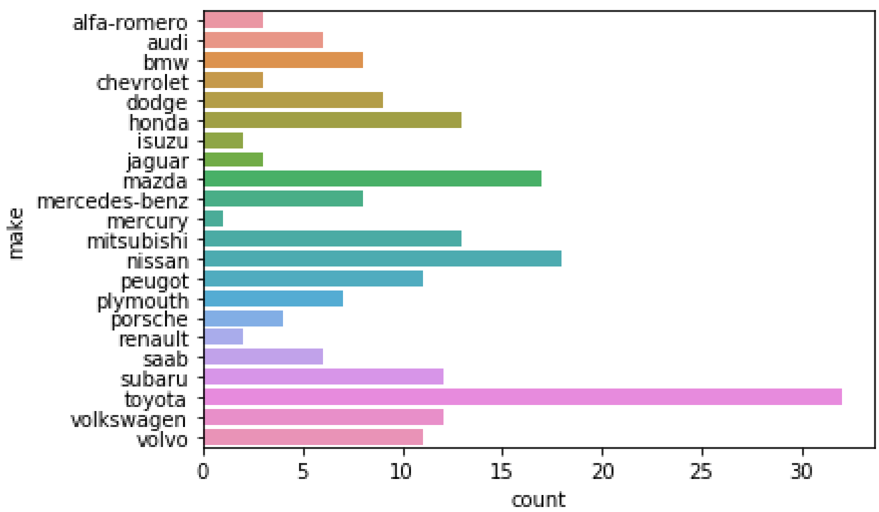
- num_of_cylinders with 4 shows the highest count
- num_of_cylinders with 3 and 12 shows minimum count

In [14]:

```
sns.countplot(y='make',data=autodv)
```

Out[14]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xbc4e948>
```



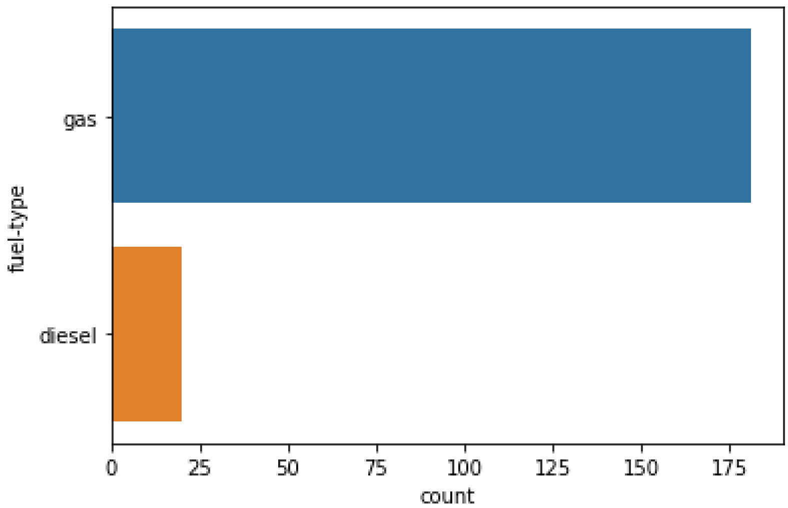
- highest 'make' from brand toyota
- lowest 'make' from brand mercury

In [16]:

```
sns.countplot(y='fuel-type',data=autodv)
```

Out[16]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xbd6e248>
```



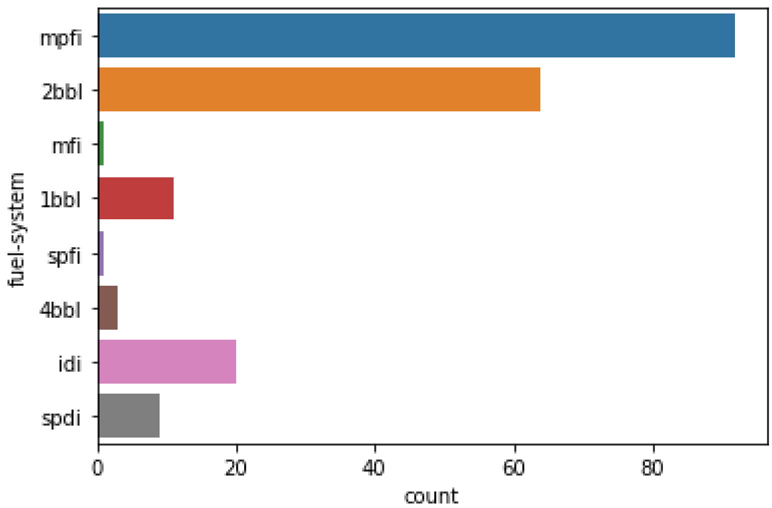
- most cars have fuel type as gas with count 175
- fewer cars have fuel type diesel which range around 20 -25

In [17]:

```
sns.countplot(y='fuel-system',data=autodv)
```

Out[17]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xbdbb648>
```



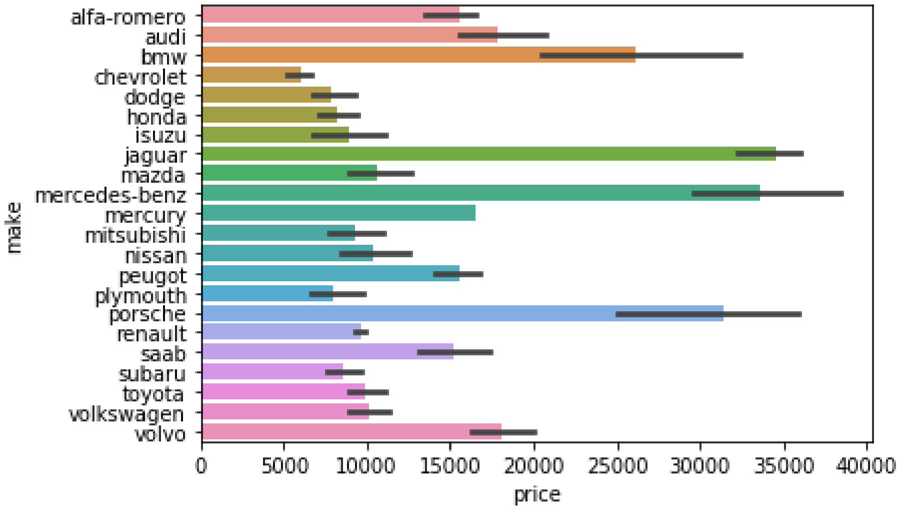
- highest 'fuel_system' is mpfi with 80+ count
- lowest 'fuel_sytem' of cars in dataset are mfi and spfi

In [18]:

```
## Bar plot
sns.barplot(x='price',y='make',data=allen)
```

Out[18]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xbe3dcc8>
```



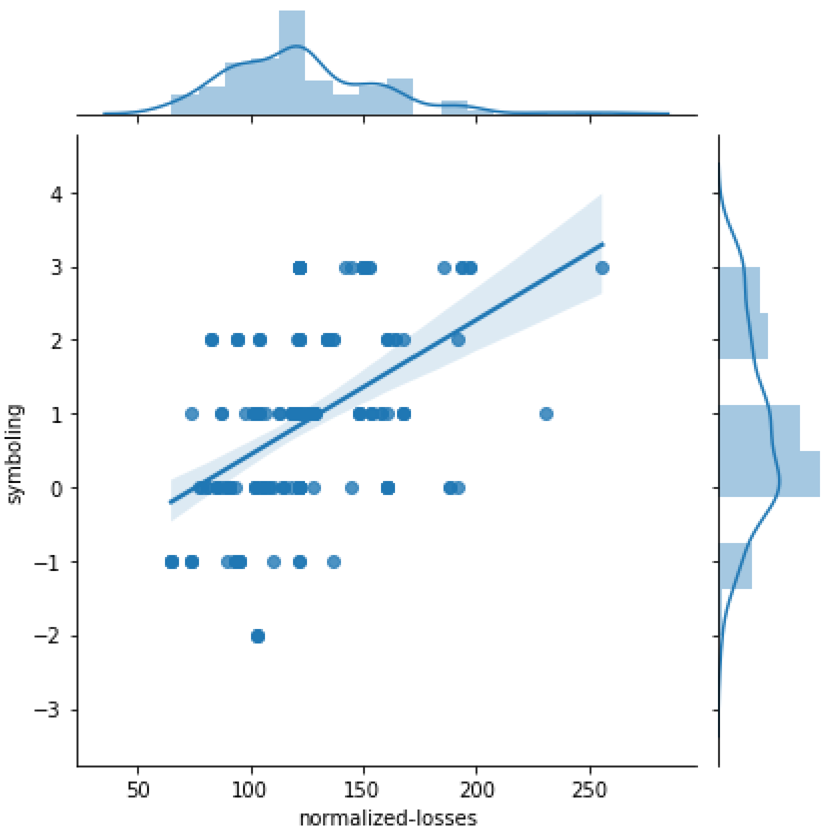
- jaguar brand shows highest price range -chevrolet brand shows lowest price range

In [21]:

```
sns.jointplot(x='normalized-losses',y='symboling',data=autodv,kind='reg')
```

Out[21]:

```
<seaborn.axisgrid.JointGrid at 0xc128848>
```



- with symboling values range from 0 to 1 , there seems to have a relation between normalized-losses and symboling

In [24]:

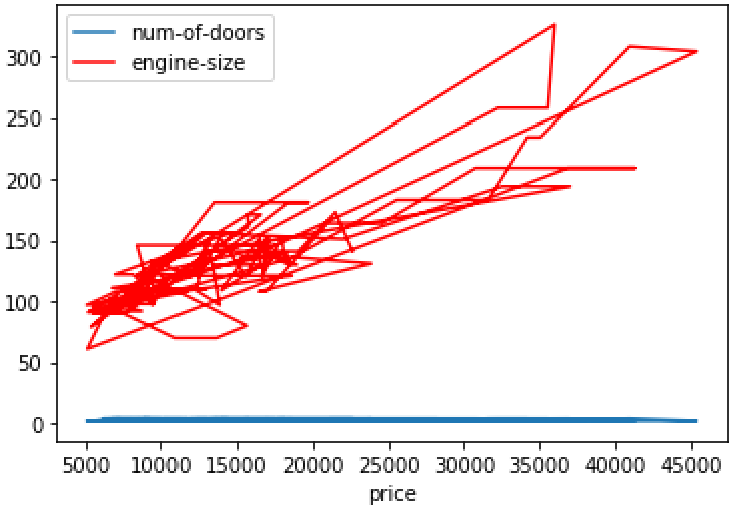
```
#line plot

ax = plt.gca()
#gca stands for get current axis

autodv.plot(kind = 'line', x = 'price', y = 'num-of-doors', ax=ax)
autodv.plot(kind = 'line', x = 'price', y = 'engine-size', color = 'red',ax=ax)
```

Out[24]:

```
<matplotlib.axes._subplots.AxesSubplot at 0xc310048>
```



- graph does not show any relationship of enginesize and noof doors with price

In [ ]:

```

```