In [1]:

```
#201600282 엄기산
```

In [5]:

```
from IPython.core.interactiveshell import InteractiveShell
InteractiveShell.ast_node_interactivity = "all"
```

In [4]:

```
import pandas
from sklearn.datasets import load_iris
iris=load_iris()
irisdf=pandas.DataFrame(iris.data,columns=iris.feature_names)
irisdf['target']=iris.target
irisdf['target']=irisdf['target'].map({0:"seatosa",1:"versicolor",2:"virginica"})
irisdf
```

Out[4]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
0	5.1	3.5	1.4	0.2	seatosa
1	4.9	3.0	1.4	0.2	seatosa
2	4.7	3.2	1.3	0.2	seatosa
3	4.6	3.1	1.5	0.2	seatosa
4	5.0	3.6	1.4	0.2	seatosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	virginica
146	6.3	2.5	5.0	1.9	virginica
147	6.5	3.0	5.2	2.0	virginica
148	6.2	3.4	5.4	2.3	virginica
149	5.9	3.0	5.1	1.8	virginica

150 rows × 5 columns

In [12]:

```
import matplotlib.pyplot as plt
SL=irisdf.iloc[:,0]
SW=irisdf.iloc[:,1]
PL=irisdf.iloc[:,2]
PW=irisdf.iloc[:,3]
name=irisdf.iloc[:,4]
plt.scatter(PL,SL,c=iris.target)
plt.title("PL / SL Eom Gi san")
plt.xlabel("petal length")
plt.ylabel("sepal length")
plt.show()
```

Out[12]:

```
<matplotlib.collections.PathCollection at 0x20a8704df70>
```

Out[12]:

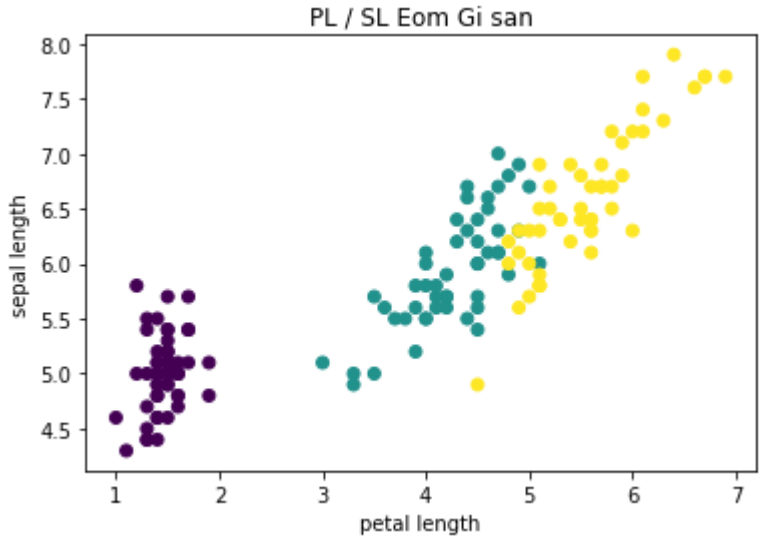
```
Text(0.5, 1.0, 'PL / SL Eom Gi san')
```

Out[12]:

```
Text(0.5, 0, 'petal length')
```

Out[12]:

```
Text(0, 0.5, 'sepal length')
```



In [50]:

```
irisav=pandas.DataFrame(iris.data,columns=iris.feature_names)
irisav['target']=iris.target
irisav['target']=irisav['target'].map({0:"seatosa",1:"versicolor",2:"virginica"})
#기존에 있는 irisdf로 평균을 구하려고 했지만.. 위의 3문장을 추가해 irisav 데이터 프레임을 새롭게 만들
#aver 마지막 열구분에 target이 안나와서 추가하였습니다. 왜 그러는지는 잘 모르겠습니다..
aver = irisav.groupby(iris.target).mean()
aver

import matplotlib.pyplot as plt

aver.T.plot.bar()
plt.xlabel("variable")
plt.ylabel("aver")
plt.ylim(0,8)
plt.show()

#target내에서 PL PW SL SW 들의 연관성에 집중하는 것이 아니므로 선형, 간단한 그래프는 적절하지 않다고
#또한 히스토그램은 한다지의 변수의 데이터를 나타내기 적절한 그래프이므로
#SW SL PL PW 4가지 변수에 대해 나타내야하는 상황에서 적합하지 않다고 생각했습니다.
#평균값이 아닌 모든 데이터를 그래프로 나타낸다면 산점도 그래프가 적절하겠지만
#이 경우에는 평균 데이터를 그래프로 나타내기 때문에 막대 그래프가 적절하다고 생각했습니다.
```

Out[50]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
0	5.006	3.428	1.462	0.246
1	5.936	2.770	4.260	1.326
2	6.588	2.974	5.552	2.026

Out[50]:

```
<AxesSubplot:>
```

Out[50]:

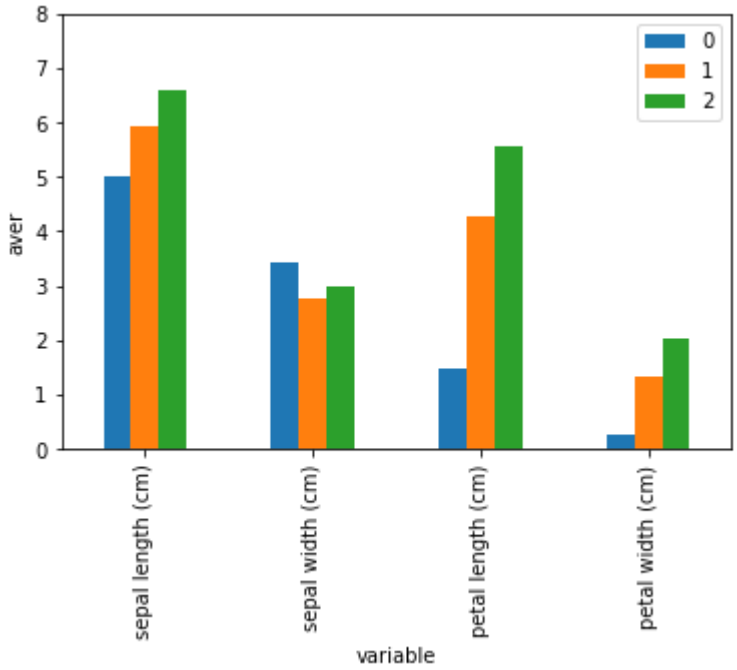
```
Text(0.5, 0, 'variable')
```

Out[50]:

```
Text(0, 0.5, 'aver')
```

Out[50]:

```
(0.0, 8.0)
```


8
7
6
5
4
3
2
1
0
aver
0
1
2
sepal length (cm)
sepal width (cm)
petal length (cm)
petal width (cm)
variable

In [79]:

```
from sklearn.model_selection import train_test_split

X_train, X_test = train_test_split(irisav, test_size=0.25)

X_train.to_csv('./X_train.csv')
X_test.to_csv('./X_test.csv')

train = pandas.read_csv("./X_train.csv",index_col = 'Unnamed: 0')
test = pandas.read_csv("./X_test.csv", index_col = 'Unnamed: 0')
#기존의 인덱스값이 첫 열에 있어 그 열을 제외하고 불러오기위해 index_col ='unnamed: 0' 명령을 추가하

aver_train = train.groupby(train.target).mean()
aver_train
aver_test = test.groupby(test.target).mean()
aver_test
```

Out[79]:

target	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
seatosa	5.060526	3.494737	1.476316	0.239474
versicolor	5.940000	2.742500	4.242500	1.310000
virginica	6.570588	2.932353	5.538235	1.976471

Out[79]:

target	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)
seatosa	4.833333	3.216667	1.416667	0.266667
versicolor	5.920000	2.880000	4.330000	1.390000
virginica	6.625000	3.062500	5.581250	2.131250

In [ ]:

```
#201600282 엄기산
```