201600779 김영민

In [1]:

```
import pandas as pd
from sklearn.datasets import load_iris
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
iris = load_iris()
df = pd.DataFrame(iris.data,columns = iris.feature_names)
```

In [3]:

```
df['Target'] = iris.target
```

In [4]:

```
dic = {}
for k,v in enumerate(iris.target_names):
    dic[k] = v
df['Target'] = df['Target'].map(dic)
```

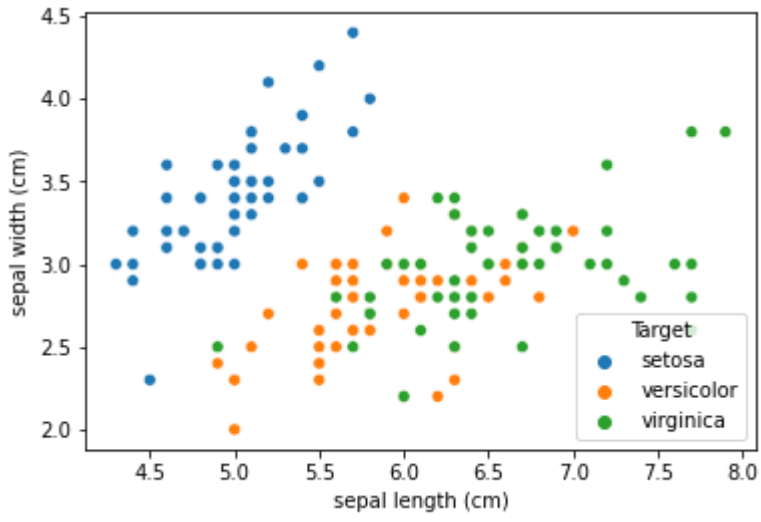
1번

In [5]:

```
sns.scatterplot(df['sepal length (cm)'],df['sepal width (cm)'],hue=df['Target'])
```

Out[5]:

<AxesSubplot:xlabel='sepal length (cm)', ylabel='sepal width (cm)'>

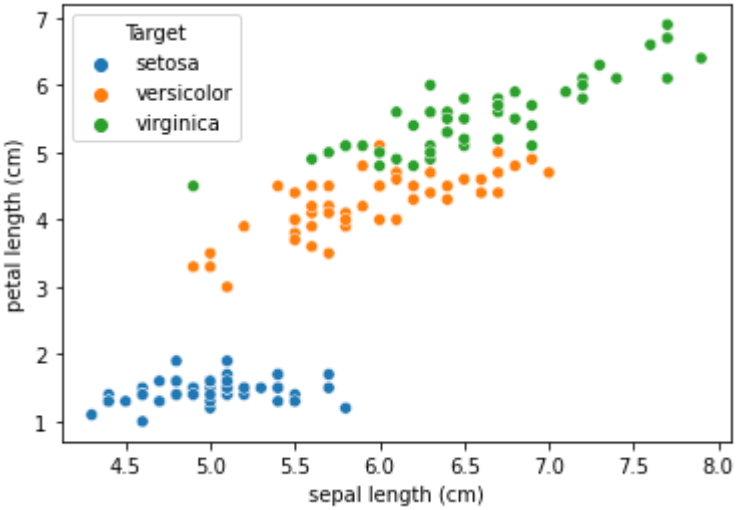


In [6]:

```
sns.scatterplot(df['sepal length (cm)'],df['petal length (cm)'],hue=df['Target'])
```

Out[6]:

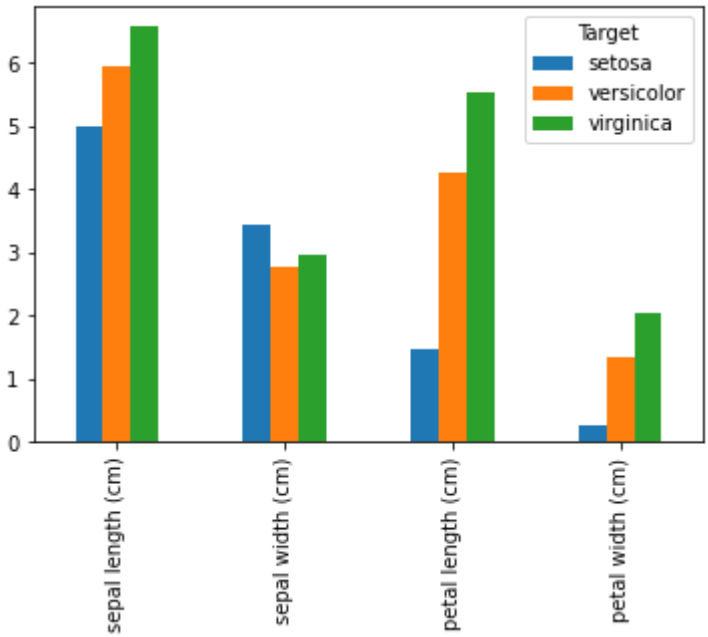
<AxesSubplot:xlabel='sepal length (cm)', ylabel='petal length (cm)'>



2번

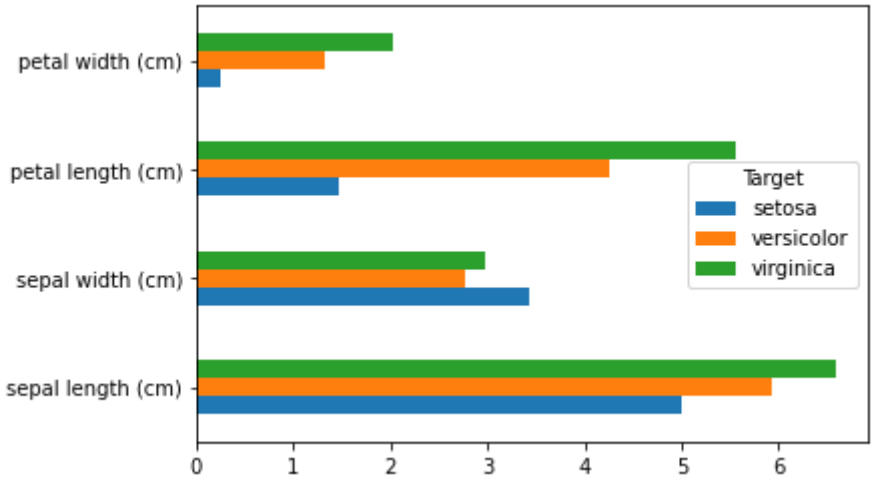
In [7]:

```
mean_df = df.groupby('Target').mean()
mean_df.T.plot.bar()
plt.show()
```



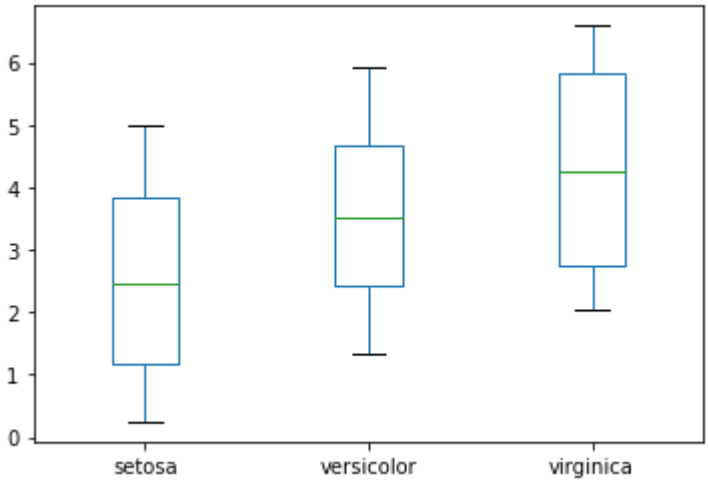
In [8]:

```
mean_df.T.plot.barh()
plt.show()
```



In [9]:

```
mean_df.T.plot.box()
plt.show()
```



barplot이 가장 좋다. Target별 분류와 평균값을 보기좋게 나타낼 수 있기 때문이다.

3번

In [10]:

```
from sklearn.model_selection import train_test_split
train,test = train_test_split(df,test_size=0.25)
```

In [11]:

```
train.to_csv('train_dataset.csv',index=False)
test.to_csv('test_dataset.csv',index=False)
```

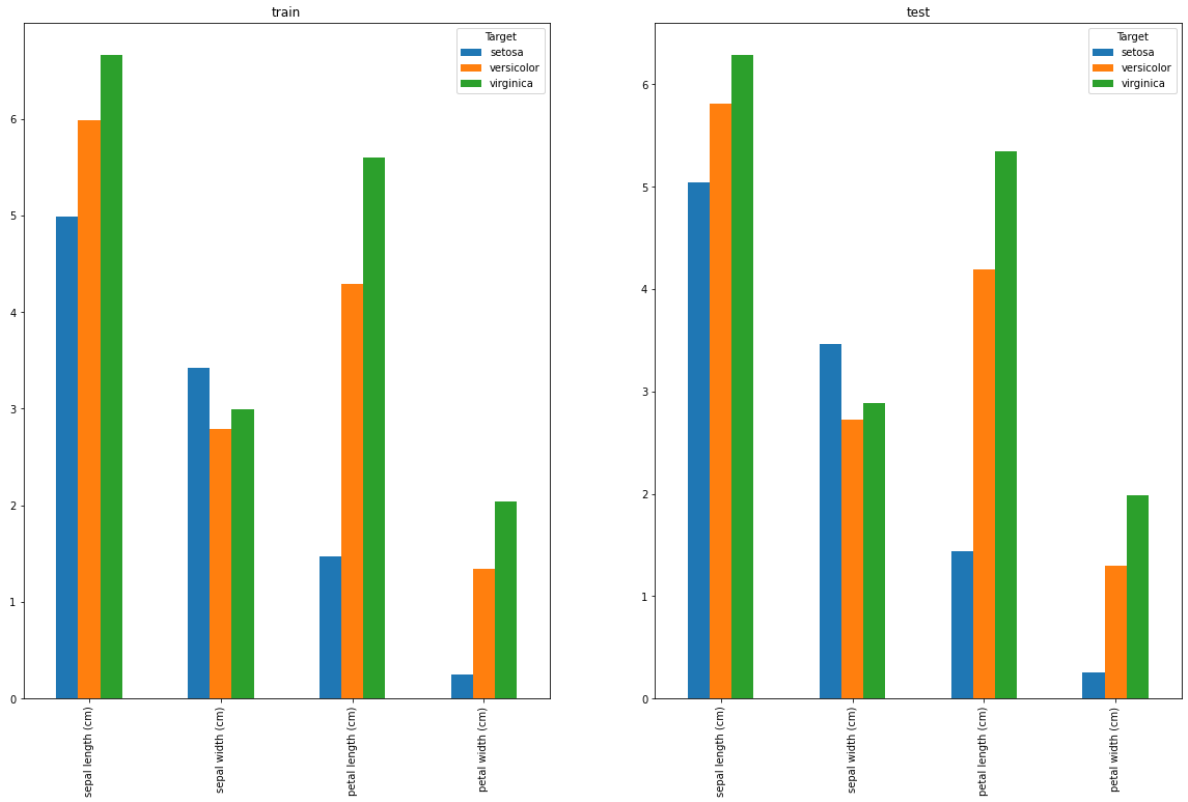
4번

In [12]:

```
train = pd.read_csv('train_dataset.csv')
test = pd.read_csv('test_dataset.csv')
```

In [13]:

```
fig,(ax1,ax2) = plt.subplots(ncols=2,figsize=(20,12))
mean_df_train = train.groupby('Target').mean()
mean_df_test = test.groupby('Target').mean()
mean_df_train.T.plot.bar(ax=ax1)
mean_df_test.T.plot.bar(ax=ax2)
ax1.set(title='train')
ax2.set(title='test')
plt.show()
```



201600779 김영민