```python
In [1]: #when numeric data use regression or when catagorical data use classification
        import pandas as pd
        from sklearn import tree
        from sklearn.tree import DecisionTreeRegressor
        from sklearn.model_selection import train_test_split,cross_val_score
        import matplotlib.pyplot as plt
```

```python
In [2]: data=pd.read_csv("Pollution.csv")
        data.head()
```

Out[2]:

	Pollution	Temp	Industry	Population	Wind	Rain	WetDays
0	24	61.5	368	497	9.1	48.34	115
1	30	55.6	291	593	8.3	43.11	123
2	56	55.9	775	622	9.5	35.89	105
3	28	51.0	137	176	8.7	15.17	89
4	14	68.4	136	529	8.8	54.47	116

```python
In [3]: data.columns
```

```
Out[3]: Index(['Pollution ', 'Temp', 'Industry', 'Population', 'Wind', 'Rain',
               'WetDays'],
              dtype='object')
```

```python
In [4]: y=data['Pollution ']
        x=data.drop(['Pollution '],axis=1)
```

```python
In [5]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.20,shuffle=True)
```

```python
In [6]: model=DecisionTreeRegressor(max_depth=3,max_leaf_nodes=10)
```

```python
In [7]: model.fit(x_train,y_train)
```

```
Out[7]: DecisionTreeRegressor(max_depth=3, max_leaf_nodes=10)
```

```python
In [8]: cross_val_score(model,x,y,cv=10)
```

```
Out[8]: array([-1.47944494e+00, -6.47547071e-01, -2.84114671e-01,  4.03194620e-02,
               -2.25434265e+00, -1.42415165e-03, -9.51222591e-02, -1.81752909e-04,
               -8.28505193e+00,  1.80157977e-01])
```

```
In [9]: fn=data.columns
        cn=['setosa',"veginica","vercicolor"]
        fig,axes=plt.subplots(nrows=1, ncols=1,dpi=300,figsize=(4,4))
        tree.plot_tree(model,feature_names=fn,class_names=cn,filled=True)
        plt.show()
```

Temp <= 748.0
mse = 447.89
samples = 32
value = 27.719

Rain <= 108.0
mse = 159.27
samples = 29
value = 22.621

Industry <= 2060.0
mse = 558.0
samples = 3
value = 77.0

Pollution <= 51.25
mse = 27.41
samples = 12
value = 13.417

Industry <= 228.5
mse = 150.339
samples = 17
value = 29.118

Temp <= 891.0
mse = 20.25
samples = 2
value = 60.5

mse = 0.0
samples = 1
value = 110.0

mse = 30.25
samples = 2
value = 22.5

mse = 7.04
samples = 10
value = 11.6

mse = 92.188
samples = 4
value = 42.25

mse = 98.84
samples = 13
value = 25.077

mse = 0.0
samples = 1
value = 56.0

mse = 0.0
samples = 1
value = 65.0