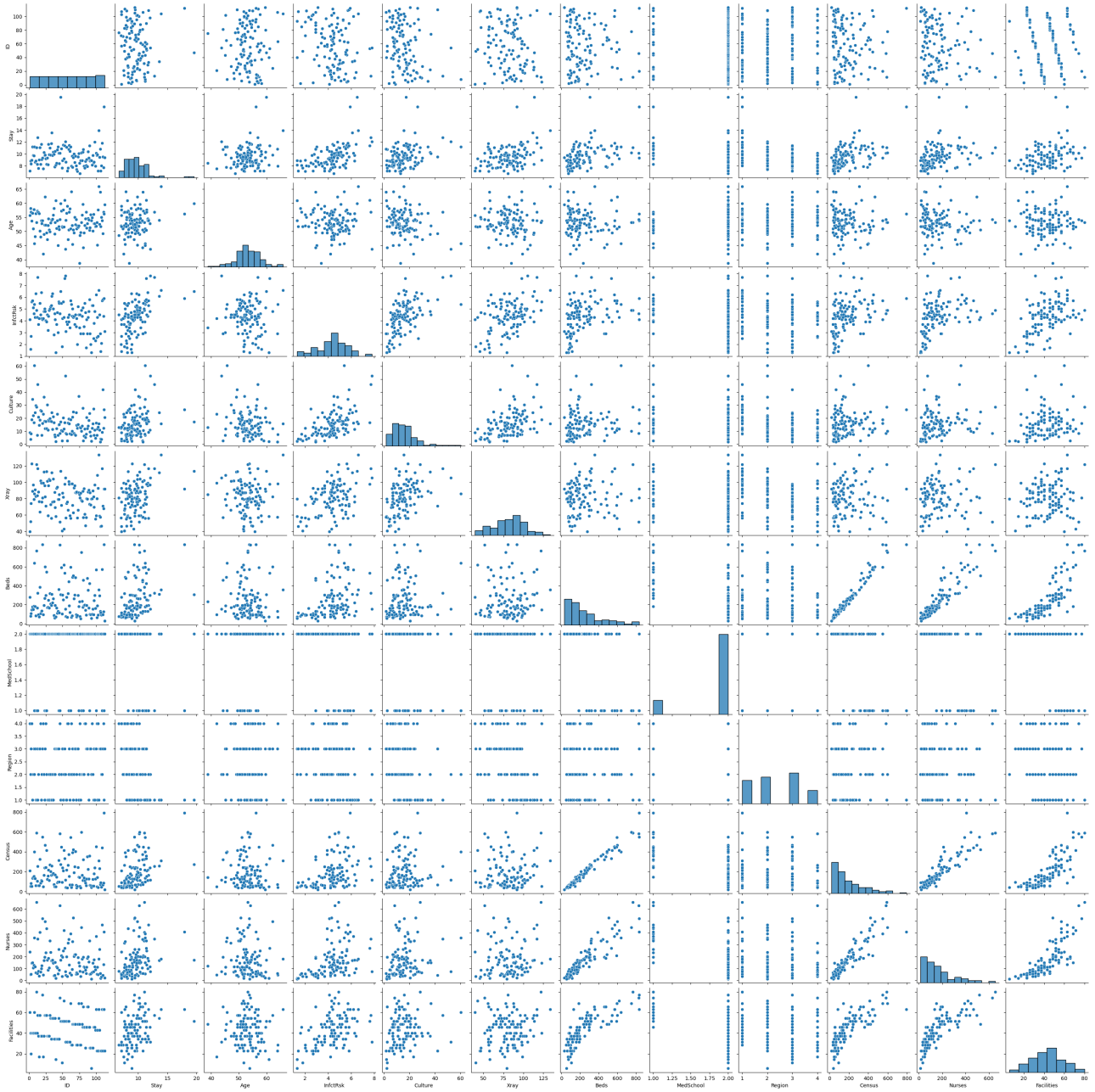
1-(1).


ID
Stay
Age
InfctRsk
Culture
Xray
Beds
MedSchool
Region
Census
Nurses
Facilities

1-(2).

```
correlation analysis
                 ID      Stay       Age  InfctRsk   Culture      Xray  \
ID         1.000000 -0.022387  0.037258 -0.211316 -0.267415 -0.166020
Stay      -0.022387  1.000000  0.188914  0.533444  0.326684  0.382482
Age        0.037258  0.188914  1.000000  0.001093 -0.225847 -0.018855
InfctRsk  -0.211316  0.533444  0.001093  1.000000  0.559159  0.453392
Culture   -0.267415  0.326684 -0.225847  0.559159  1.000000  0.424962
Xray      -0.166020  0.382482 -0.018855  0.453392  0.424962  1.000000
Beds      -0.035653  0.409265 -0.058823  0.359770  0.139725  0.045820
MedSchool  0.006071 -0.296951  0.145126 -0.233030 -0.242744 -0.086697
Region     0.101505 -0.492130 -0.020432 -0.192281 -0.308278 -0.296344
Census    -0.027056  0.473885 -0.054775  0.381411  0.142948  0.062914
Nurses    -0.133531  0.340367 -0.082945  0.393981  0.198900  0.077381
Facilities -0.097857  0.355538 -0.040451  0.412601  0.185131  0.111928

                Beds  MedSchool    Region    Census    Nurses  Facilities
ID         -0.035653   0.006071  0.101505 -0.027056 -0.133531   -0.097857
Stay        0.409265  -0.296951 -0.492130  0.473885  0.340367    0.355538
Age        -0.058823   0.145126 -0.020432 -0.054775 -0.082945   -0.040451
InfctRsk    0.359770  -0.233030 -0.192281  0.381411  0.393981    0.412601
Culture     0.139725  -0.242744 -0.308278  0.142948  0.198900    0.185131
Xray        0.045820  -0.086697 -0.296344  0.062914  0.077381    0.111928
Beds        1.000000  -0.591180 -0.105627  0.980998  0.915504    0.794524
MedSchool  -0.591180   1.000000  0.102668 -0.614757 -0.588240   -0.524390
Region     -0.105627   0.102668  1.000000 -0.152744 -0.112681   -0.211532
Census      0.980998  -0.614757 -0.152744  1.000000  0.907897    0.778063
Nurses      0.915504  -0.588240 -0.112681  0.907897  1.000000    0.783505
Facilities  0.794524  -0.524390 -0.211532  0.778063  0.783505    1.000000
```

1-(3)-1.

```
ID            0
Stay          0
Age           0
InfctRsk      0
Culture       0
Xray          0
Beds          0
MedSchool     0
Region        0
Census        0
Nurses        0
Facilities    0
dtype: int64
```

1-(3)-2.

결측치가 없으므로 전처리가 필요 없다.

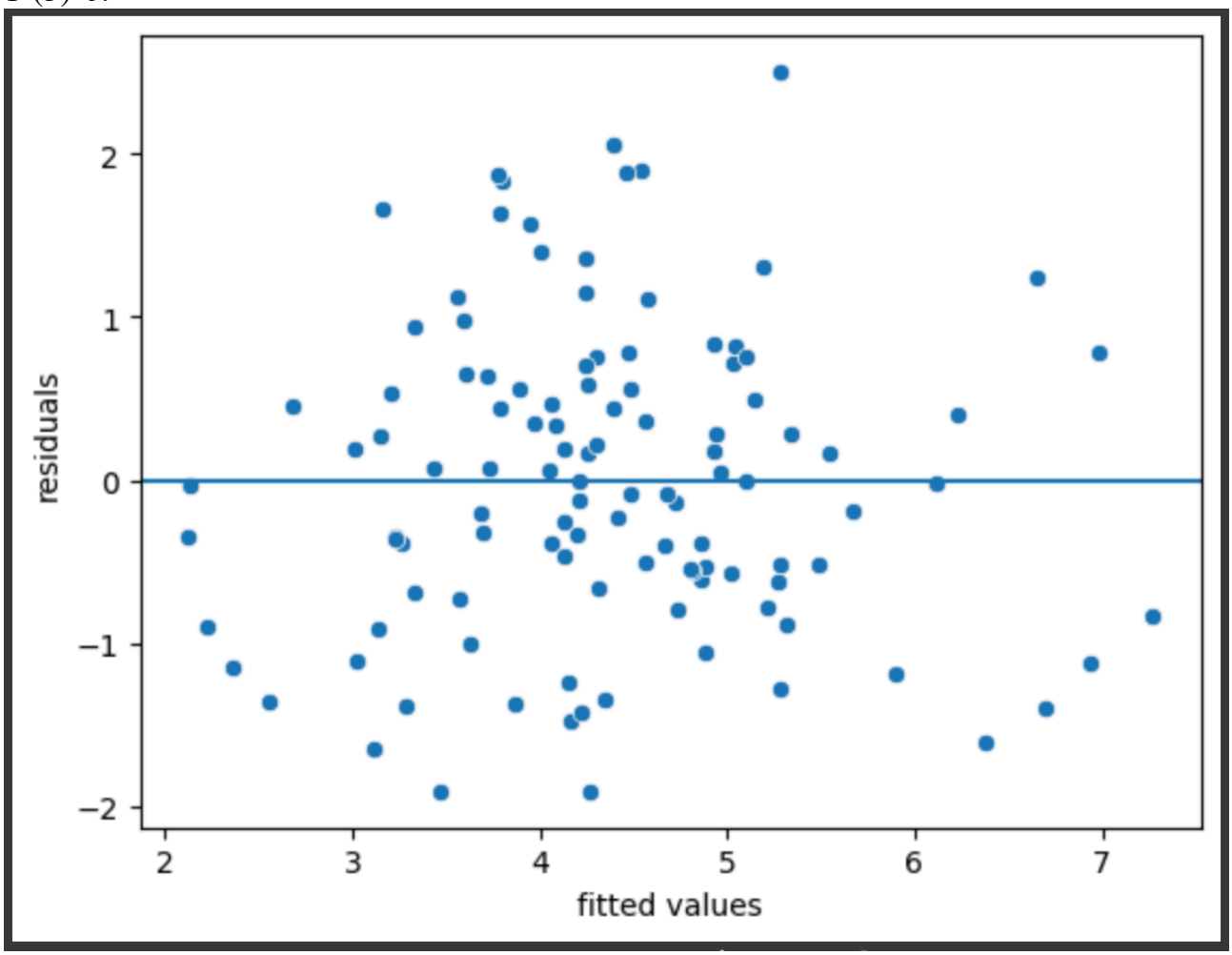
1-(3)-4.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-3.0366	1.447	-2.098	0.038	-5.908	-0.166
ID	-0.0020	0.003	-0.697	0.488	-0.008	0.004
Stay	0.2453	0.071	3.456	0.001	0.104	0.386
Age	0.0095	0.022	0.441	0.660	-0.033	0.052
Culture	0.0520	0.011	4.818	0.000	0.031	0.073
Xray	0.0124	0.005	2.333	0.022	0.002	0.023
Beds	-0.0031	0.003	-1.162	0.248	-0.008	0.002
MedSchool	0.5428	0.322	1.687	0.095	-0.096	1.181
Region	0.3006	0.105	2.867	0.005	0.093	0.509
Census	0.0029	0.003	0.848	0.398	-0.004	0.010
Nurses	0.0018	0.002	1.056	0.294	-0.002	0.005
Facilities	0.0230	0.010	2.290	0.024	0.003	0.043

1-(3)-5.

```
mae: 0.5930129350985645
mse: 0.5605675300605998
rmse: 0.7487105783015222
r2: 0.5998431660026755
```

1-(3)-6.



1-(3)-7.

mae, mse, rmse, r-square의 점수를 바탕으로 살펴본 결과 기대치는 준수한 편이라고 할 수 있다. mse, rmse 스코어를 보았을 때, 이상치가 꽤나 존재한다는 점을 알 수 있고, r-square 점수를 보았을 때, 전반적으로 성능이 썩 좋다고 볼수 없다는 것을 알 수 있다.

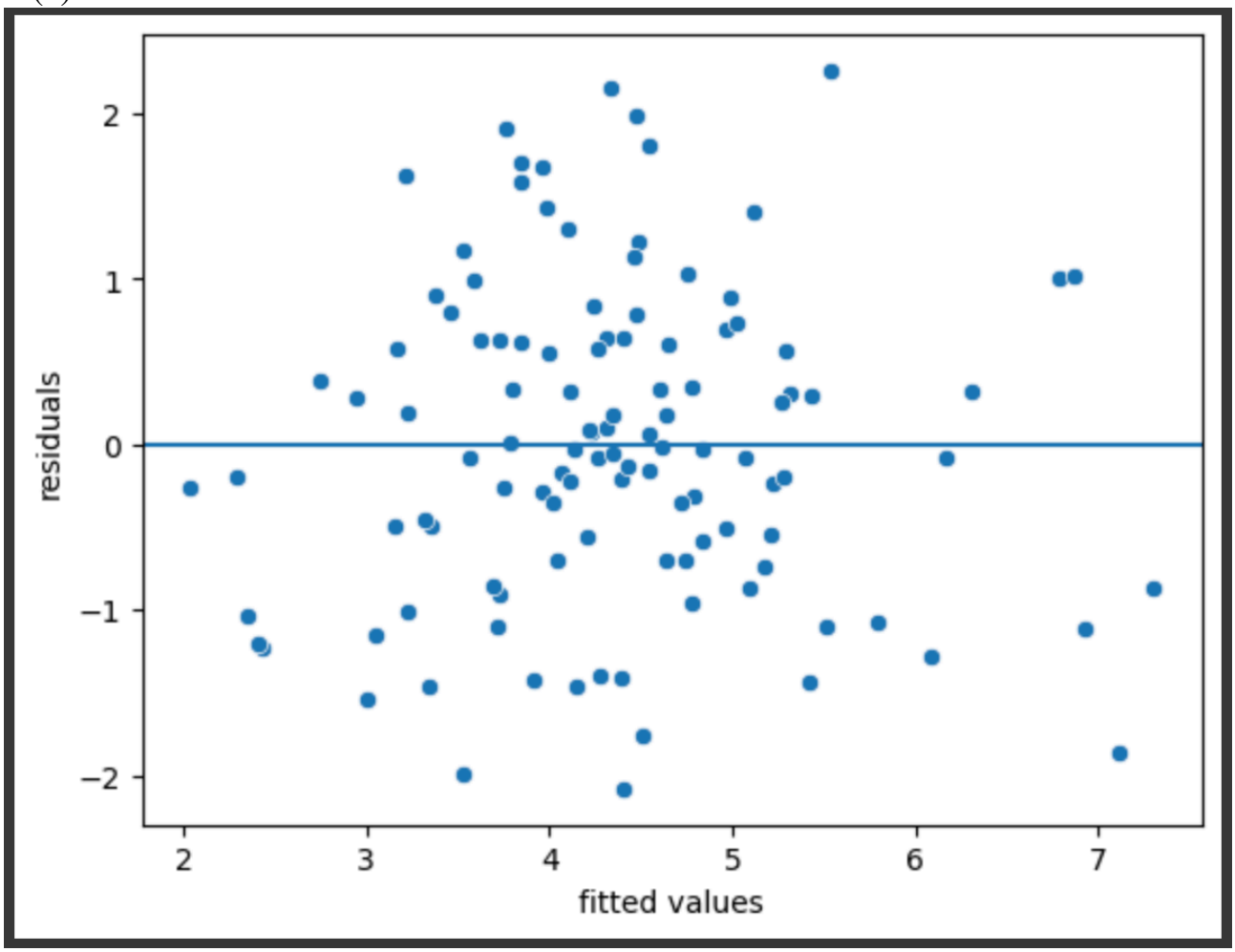
1-(4)-1.

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-2.6944	1.067	-2.526	0.013	-4.809	-0.580
Stay	0.2601	0.058	4.513	0.000	0.146	0.374
Culture	0.0528	0.010	5.376	0.000	0.033	0.072
Xray	0.0126	0.005	2.435	0.017	0.002	0.023
MedSchool	0.4631	0.291	1.589	0.115	-0.115	1.041
Region	0.2904	0.101	2.871	0.005	0.090	0.491
Facilities	0.0262	0.007	3.781	0.000	0.012	0.040

1-(4)-2.

```
mae: 0.5691602303972083
mse: 0.5376815255396669
rmse: 0.7332677038706034
r2: 0.6161801648847816
```

1-(4)-3.



1-(4)-4.

mae, mse, rmse, r-square의 점수를 바탕으로 살펴본 결과 기대치는 준수한 편이라고 할 수 있다. mse, rmse 스코어를 보았을 때, 이상치가 꽤나 존재한다는 점을 알 수 있고, r-square 점수를 보았을 때, 1-(3)번과 비교하여 전반적으로 약간 나아진 성능을 보이지만, 눈에 띄게 좋아졌다고 보기는 어렵다.

1-(5).

회귀계수가 0.01 이하인 변수를 제외하고 학습을 수행하였는데, 회귀계수가 0.1 이상인 변수가 존재하기 때문에 변수 제외 회귀계수의 기준치를 조금 더 높인 후에 학습을 진행하면 조금 더 좋은 결과가 나올 것으로 보인다.

2-1.

```
Sex          0
Height       0
HandSpan     0
dtype: int64
```

2-2.

결측치가 없으므로 전처리가 필요 없다.

2-3.

```
[[0.65461268 1.15634789]]
[-68.84155381]
```

2-4.

```
0.9411764705882353
```

2-6.

```
[[18  0]
 [ 2 14]]
정확도: 0.9412, 정밀도: 0.9000, 재현율: 1.0000, F1:0.9474
ROC AUC 값: 1.0000
```

2-7.

각각의 독립변수에 대한 절편의 값이 상당히 높고, 독립변수의 갯수가 상대적으로 적고, 체형과 성별의 상관관계가 눈에 띄게 크기 때문에 아주 높은 성능과 정확도를 보여주고있다.

3-1.

```
mpg             0
cylinders       0
displacement    0
horsepower      6
weight          0
acceleration    0
model_year      0
origin          0
name            0
dtype: int64
```

3-2.

결측치가 존재하므로 전처리 수행

3-5.

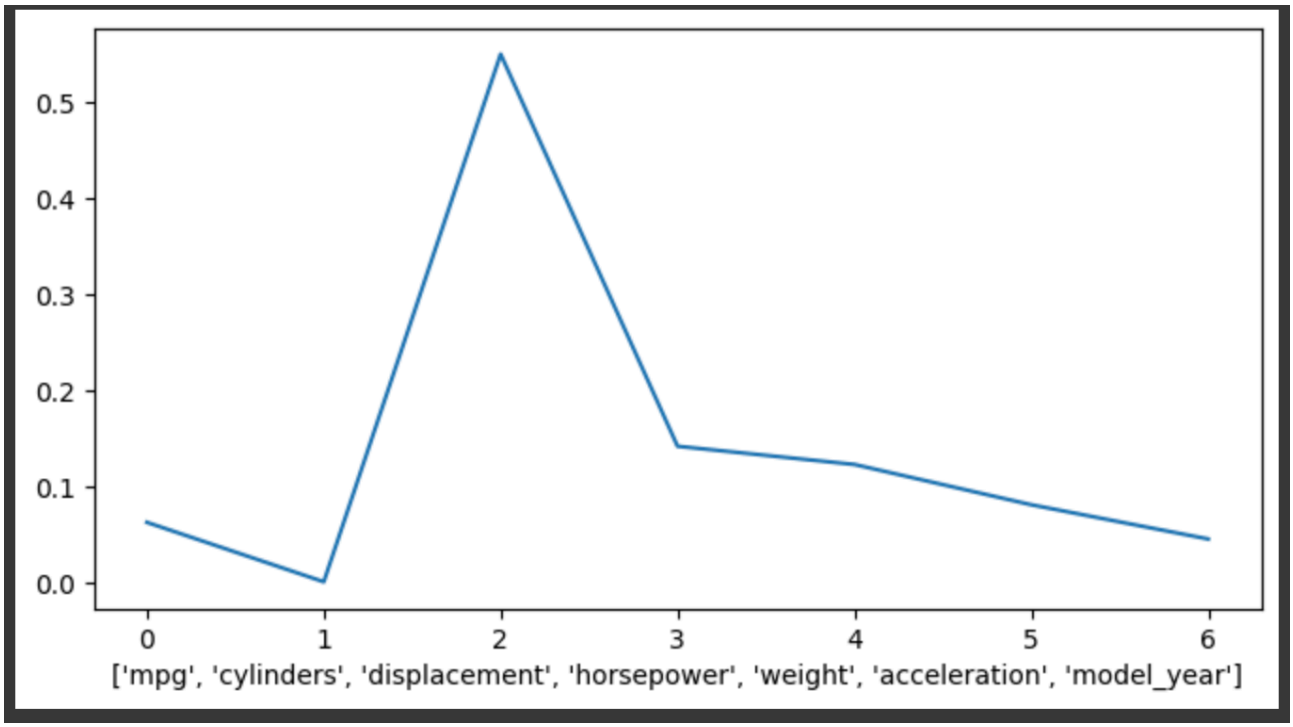
```
0.80
```

3-6, 3-7, 3-8.

```
best parameters :  {'max_depth': 3, 'min_samples_split': 3}
best score :  0.7508547008547009
0.7721518987341772
best parameters :  {'max_depth': 5, 'min_samples_split': 3}
best score :  0.7956654456654456
0.8227848101265823
best parameters :  {'max_depth': 4, 'min_samples_split': 3}
best score :  0.7765262515262515
0.8481012658227848
best parameters :  {'max_depth': 5, 'min_samples_split': 5}
best score :  0.7956959706959706
0.8227848101265823
best parameters :  {'max_depth': 4, 'min_samples_split': 2}
best score :  0.7733821733821733
0.8481012658227848
```

3-9.

```
[0.06194744 0.         0.54972301 0.14122994 0.12223345 0.08025846
 0.0446077 ]
```



중요도가 높은 상위 5개 피처는 displacement, horsepower, weight, acceleration, mpg이다.

3-10. (tree.dot 결과파일 첨부)

```
digraph Tree {
node [shape=box, style="filled", color="black", fontname="helvetica"] ;
edge [fontname="helvetica"] ;
0 [label="displacement <= 134.5\ngini = 0.558\nsamples = 313\nvalue = [54, 71, 188]\nclass = japan", fillcolor="#c29ff2"] ;
1 [label="mpg <= 30.95\ngini = 0.623\nsamples = 136\nvalue = [48, 64, 24]\nclass = europe", fillcolor="#dbfae8"] ;
0 -> 1 [labeldistance=2.5, labelangle=45, headlabel="True"] ;
2 [label="horsepower <= 87.0\ngini = 0.622\nsamples = 74\nvalue = [36, 24, 14]\nclass = usa", fillcolor="#f9e1cf"] ;
1 -> 2 ;
3 [label="displacement <= 97.25\ngini = 0.578\nsamples = 38\nvalue = [21, 5, 12]\nclass = usa", fillcolor="#f6d3ba"] ;
2 -> 3 ;
4 [label="model_year <= 74.5\ngini = 0.43\nsamples = 22\nvalue = [16, 4, 2]\nclass = usa", fillcolor="#eeab7b"] ;
3 -> 4 ;
5 [label="acceleration <= 15.0\ngini = 0.153\nsamples = 12\nvalue = [11, 0, 1]\nclass = usa", fillcolor="#e78c4b"] ;
4 -> 5 ;
6 [label="horsepower <= 75.5\ngini = 0.444\nsamples = 3\nvalue = [2, 0, 1]\nclass = usa", fillcolor="#f2c09c"] ;
5 -> 6 ;
7 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 1]\nclass = japan", fillcolor="#8139e5"] ;
6 -> 7 ;
8 [label="gini = 0.0\nsamples = 2\nvalue = [2, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
6 -> 8 ;
9 [label="gini = 0.0\nsamples = 9\nvalue = [9, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
5 -> 9 ;
10 [label="acceleration <= 15.65\ngini = 0.58\nsamples = 10\nvalue = [5, 4, 1]\nclass = usa", fillcolor="#fbeade"] ;
4 -> 10 ;
11 [label="gini = 0.0\nsamples = 4\nvalue = [4, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
10 -> 11 ;
12 [label="displacement <= 93.5\ngini = 0.5\nsamples = 6\nvalue = [1, 4, 1]\nclass = europe", fillcolor="#88efb3"] ;
10 -> 12 ;
13 [label="acceleration <= 19.35\ngini = 0.5\nsamples = 2\nvalue = [1, 0, 1]\nclass = usa", fillcolor="#ffffff"] ;
12 -> 13 ;
14 [label="gini = 0.0\nsamples = 1\nvalue = [1, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
13 -> 14 ;
15 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 1]\nclass = japan", fillcolor="#8139e5"] ;
13 -> 15 ;
16 [label="gini = 0.0\nsamples = 4\nvalue = [0, 4, 0]\nclass = europe", fillcolor="#39e581"] ;
12 -> 16 ;
17 [label="weight <= 2429.5\ngini = 0.508\nsamples = 16\nvalue = [5, 1, 10]\nclass = japan", fillcolor="#c6a5f3"] ;
3 -> 17 ;
18 [label="weight <= 2219.5\ngini = 0.439\nsamples = 14\nvalue = [3, 1, 10]\nclass = japan", fillcolor="#af81ee"] ;
17 -> 18 ;
19 [label="displacement <= 99.5\ngini = 0.594\nsamples = 8\nvalue = [3, 1, 4]\nclass = japan", fillcolor="#e6d7fa"] ;
18 -> 19 ;
20 [label="horsepower <= 81.5\ngini = 0.5\nsamples = 6\nvalue = [1, 1, 4]\nclass = japan", fillcolor="#b388ef"] ;
19 -> 20 ;
21 [label="model_year <= 77.5\ngini = 0.32\nsamples = 5\nvalue = [0, 1, 4]\nclass = japan", fillcolor="#a06aec"] ;
20 -> 21 ;
22 [label="gini = 0.0\nsamples = 3\nvalue = [0, 0, 3]\nclass = japan", fillcolor="#8139e5"] ;
21 -> 22 ;
23 [label="weight <= 2145.0\ngini = 0.5\nsamples = 2\nvalue = [0, 1, 1]\nclass = europe", fillcolor="#ffffff"] ;
21 -> 23 ;
24 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0]\nclass = europe", fillcolor="#39e581"] ;
23 -> 24 ;
25 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 1]\nclass = japan", fillcolor="#8139e5"] ;
23 -> 25 ;
26 [label="gini = 0.0\nsamples = 1\nvalue = [1, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
20 -> 26 ;
27 [label="gini = 0.0\nsamples = 2\nvalue = [2, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
19 -> 27 ;
28 [label="gini = 0.0\nsamples = 6\nvalue = [0, 0, 6]\nclass = japan", fillcolor="#8139e5"] ;
18 -> 28 ;
29 [label="gini = 0.0\nsamples = 2\nvalue = [2, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
17 -> 29 ;
30 [label="weight <= 2757.5\ngini = 0.545\nsamples = 36\nvalue = [15, 19, 2]\nclass = europe", fillcolor="#d9fae7"] ;
2 -> 30 ;
31 [label="model_year <= 81.0\ngini = 0.49\nsamples = 29\nvalue = [8, 19, 2]\nclass = europe", fillcolor="#97f1bd"] ;
30 -> 31 ;
32 [label="displacement <= 97.5\ngini = 0.417\nsamples = 27\nvalue = [8, 19, 0]\nclass = europe", fillcolor="#8cf0b6"] ;
31 -> 32 ;
33 [label="gini = 0.0\nsamples = 8\nvalue = [0, 8, 0]\nclass = europe", fillcolor="#39e581"] ;
32 -> 33 ;
34 [label="model_year <= 73.5\ngini = 0.488\nsamples = 19\nvalue = [8, 11, 0]\nclass = europe", fillcolor="#c9f8dd"] ;
32 -> 34 ;
35 [label="horsepower <= 93.0\ngini = 0.463\nsamples = 11\nvalue = [7, 4, 0]\nclass = usa", fillcolor="#f4c9aa"] ;
34 -> 35 ;
36 [label="gini = 0.0\nsamples = 4\nvalue = [4, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
35 -> 36 ;
37 [label="horsepower <= 103.5\ngini = 0.49\nsamples = 7\nvalue = [3, 4, 0]\nclass = europe", fillcolor="#cef8e0"] ;
35 -> 37 ;
38 [label="displacement <= 108.5\ngini = 0.32\nsamples = 5\nvalue = [1, 4, 0]\nclass = europe", fillcolor="#6aeca0"] ;
```

```
37 -> 38 ;
39 [label="gini = 0.0\nsamples = 1\nvalue = [1, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
38 -> 39 ;
40 [label="gini = 0.0\nsamples = 4\nvalue = [0, 4, 0]\nclass = europe", fillcolor="#39e581"] ;
38 -> 40 ;
41 [label="gini = 0.0\nsamples = 2\nvalue = [2, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
37 -> 41 ;
42 [label="mpg <= 23.45\ngini = 0.219\nsamples = 8\nvalue = [1, 7, 0]\nclass = europe", fillcolor="#55e993"] ;
34 -> 42 ;
43 [label="gini = 0.0\nsamples = 1\nvalue = [1, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
42 -> 43 ;
44 [label="gini = 0.0\nsamples = 7\nvalue = [0, 7, 0]\nclass = europe", fillcolor="#39e581"] ;
42 -> 44 ;
45 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 2]\nclass = japan", fillcolor="#8139e5"] ;
31 -> 45 ;
46 [label="gini = 0.0\nsamples = 7\nvalue = [7, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
30 -> 46 ;
47 [label="displacement <= 97.5\ngini = 0.52\nsamples = 62\nvalue = [12, 40, 10]\nclass = europe", fillcolor="#90f0b8"] ;
1 -> 47 ;
48 [label="acceleration <= 21.25\ngini = 0.361\nsamples = 38\nvalue = [9, 29, 0]\nclass = europe", fillcolor="#76eda8"] ;
47 -> 48 ;
49 [label="horsepower <= 70.5\ngini = 0.284\nsamples = 35\nvalue = [6, 29, 0]\nclass = europe", fillcolor="#62ea9b"] ;
48 -> 49 ;
50 [label="weight <= 2120.0\ngini = 0.17\nsamples = 32\nvalue = [3, 29, 0]\nclass = europe", fillcolor="#4de88e"] ;
49 -> 50 ;
51 [label="displacement <= 80.0\ngini = 0.124\nsamples = 30\nvalue = [2, 28, 0]\nclass = europe", fillcolor="#47e78a"] ;
50 -> 51 ;
52 [label="acceleration <= 18.8\ngini = 0.408\nsamples = 7\nvalue = [2, 5, 0]\nclass = europe", fillcolor="#88efb3"] ;
51 -> 52 ;
53 [label="weight <= 1790.0\ngini = 0.444\nsamples = 3\nvalue = [2, 1, 0]\nclass = usa", fillcolor="#f2c09c"] ;
52 -> 53 ;
54 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0]\nclass = europe", fillcolor="#39e581"] ;
53 -> 54 ;
55 [label="gini = 0.0\nsamples = 2\nvalue = [2, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
53 -> 55 ;
56 [label="gini = 0.0\nsamples = 4\nvalue = [0, 4, 0]\nclass = europe", fillcolor="#39e581"] ;
52 -> 56 ;
57 [label="gini = 0.0\nsamples = 23\nvalue = [0, 23, 0]\nclass = europe", fillcolor="#39e581"] ;
51 -> 57 ;
58 [label="mpg <= 35.55\ngini = 0.5\nsamples = 2\nvalue = [1, 1, 0]\nclass = usa", fillcolor="#ffffff"] ;
50 -> 58 ;
59 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0]\nclass = europe", fillcolor="#39e581"] ;
58 -> 59 ;
60 [label="gini = 0.0\nsamples = 1\nvalue = [1, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
58 -> 60 ;
61 [label="gini = 0.0\nsamples = 3\nvalue = [3, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
49 -> 61 ;
62 [label="gini = 0.0\nsamples = 3\nvalue = [3, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
48 -> 62 ;
63 [label="displacement <= 106.0\ngini = 0.601\nsamples = 24\nvalue = [3, 11, 10]\nclass = europe", fillcolor="#f1fdf6"] ;
47 -> 63 ;
64 [label="acceleration <= 17.35\ngini = 0.34\nsamples = 10\nvalue = [1, 1, 8]\nclass = japan", fillcolor="#9d65eb"] ;
63 -> 64 ;
65 [label="acceleration <= 14.3\ngini = 0.198\nsamples = 9\nvalue = [1, 0, 8]\nclass = japan", fillcolor="#9152e8"] ;
64 -> 65 ;
66 [label="weight <= 2195.0\ngini = 0.5\nsamples = 2\nvalue = [1, 0, 1]\nclass = usa", fillcolor="#ffffff"] ;
65 -> 66 ;
67 [label="gini = 0.0\nsamples = 1\nvalue = [1, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
66 -> 67 ;
68 [label="gini = 0.0\nsamples = 1\nvalue = [0, 0, 1]\nclass = japan", fillcolor="#8139e5"] ;
66 -> 68 ;
69 [label="gini = 0.0\nsamples = 7\nvalue = [0, 0, 7]\nclass = japan", fillcolor="#8139e5"] ;
65 -> 69 ;
70 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0]\nclass = europe", fillcolor="#39e581"] ;
64 -> 70 ;
71 [label="displacement <= 120.5\ngini = 0.449\nsamples = 14\nvalue = [2, 10, 2]\nclass = europe", fillcolor="#7beeab"] ;
63 -> 71 ;
72 [label="acceleration <= 17.75\ngini = 0.278\nsamples = 12\nvalue = [0, 10, 2]\nclass = europe", fillcolor="#61ea9a"] ;
71 -> 72 ;
73 [label="gini = 0.0\nsamples = 9\nvalue = [0, 9, 0]\nclass = europe", fillcolor="#39e581"] ;
72 -> 73 ;
74 [label="displacement <= 119.5\ngini = 0.444\nsamples = 3\nvalue = [0, 1, 2]\nclass = japan", fillcolor="#c09cf2"] ;
72 -> 74 ;
75 [label="gini = 0.0\nsamples = 2\nvalue = [0, 0, 2]\nclass = japan", fillcolor="#8139e5"] ;
74 -> 75 ;
76 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0]\nclass = europe", fillcolor="#39e581"] ;
74 -> 76 ;
77 [label="gini = 0.0\nsamples = 2\nvalue = [2, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
71 -> 77 ;
78 [label="displacement <= 190.5\ngini = 0.139\nsamples = 177\nvalue = [6, 7, 164]\nclass = japan", fillcolor="#8b48e7"] ;
0 -> 78 [labeldistance=2.5, labelangle=-45, headlabel="False"] ;
79 [label="weight <= 3045.0\ngini = 0.491\nsamples = 40\nvalue = [6, 7, 27]\nclass = japan", fillcolor="#b387ef"] ;
78 -> 79 ;
80 [label="horsepower <= 115.5\ngini = 0.327\nsamples = 34\nvalue = [0, 7, 27]\nclass = japan", fillcolor="#a26cec"] ;
79 -> 80 ;
81 [label="mpg <= 19.5\ngini = 0.18\nsamples = 30\nvalue = [0, 3, 27]\nclass = japan", fillcolor="#8f4fe8"] ;
```

```
80 -> 81 ;
82 [label="gini = 0.0\nsamples = 1\nvalue = [0, 1, 0]\nclass = europe", fillcolor="#39e581"] ;
81 -> 82 ;
83 [label="horsepower <= 94.0\ngini = 0.128\nsamples = 29\nvalue = [0, 2, 27]\nclass = japan", fillcolor="#8a48e7"] ;
81 -> 83 ;
84 [label="gini = 0.0\nsamples = 21\nvalue = [0, 0, 21]\nclass = japan", fillcolor="#8139e5"] ;
83 -> 84 ;
85 [label="horsepower <= 101.0\ngini = 0.375\nsamples = 8\nvalue = [0, 2, 6]\nclass = japan", fillcolor="#ab7bee"] ;
83 -> 85 ;
86 [label="gini = 0.0\nsamples = 2\nvalue = [0, 2, 0]\nclass = europe", fillcolor="#39e581"] ;
85 -> 86 ;
87 [label="gini = 0.0\nsamples = 6\nvalue = [0, 0, 6]\nclass = japan", fillcolor="#8139e5"] ;
85 -> 87 ;
88 [label="gini = 0.0\nsamples = 4\nvalue = [0, 4, 0]\nclass = europe", fillcolor="#39e581"] ;
80 -> 88 ;
89 [label="gini = 0.0\nsamples = 6\nvalue = [6, 0, 0]\nclass = usa", fillcolor="#e58139"] ;
79 -> 89 ;
90 [label="gini = 0.0\nsamples = 137\nvalue = [0, 0, 137]\nclass = japan", fillcolor="#8139e5"] ;
78 -> 90 ;
}
```

3-11.

차량의 생산지를 분류하는데 필요한 변수들은 3~5개가 적정한 것으로 보이고, 의사 결정 나무 모델의 정확도는 다소 높은 편이나, 만족스러운 유의수준 5%이내의 만족스러운 결과를 보여주지는 못하였다.

4.

```
[[0.         0.7568534  0.76430274 0.81714342 0.76688518]
 [0.88418752 0.92630115 0.95699245 0.99553039 1.        ]]
```

5.

최소-최대 정규화와 Z-점수 정규화의 차이점은 값이 생성되는 범위다. 최소-최대 정규화는 특정 범위에 속하는 값을 생성하는 반면, Z-점수 정규화는 평균이 0이고 표준 편차가 1인 값을 생성한다. 또 다른 차이점은 이상치를 처리하는 방식이다. 최소-최대 정규화는 이상치가 값 범위를 왜곡할 수 있으므로 Z-점수 정규화만큼 이상치에 강하지 않다.

6.

```
   color_blue  color_red
0           0          1
1           1          0
2           1          0
   blue  red
0     0    1
1     1    0
2     1    0
   color_blue  color_red
0           0          1
1           1          0
2           1          0
   color_blue  color_red
0           0          1
1           1          0
2           1          0
```

7.

```
first_name        1
last_name         1
age               1
sex               1
preTestScore      2
postTestScore     2
dtype: int64
```

8.

```
entropy_outlook: 1.5774062828523454
entropy_temperature: 1.5566567074628228
entropy_humidity: 1.0
entropy_windy: 0.9852281360342515
entropy_feature: 0.9402859586706311
```