In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
import seaborn as sns
```

In [2]:
```
df_merc = pd.read_csv("C:/Users/user/Desktop/My learning/ClinSoft/merc.csv
df_merc
```

Out[2]:

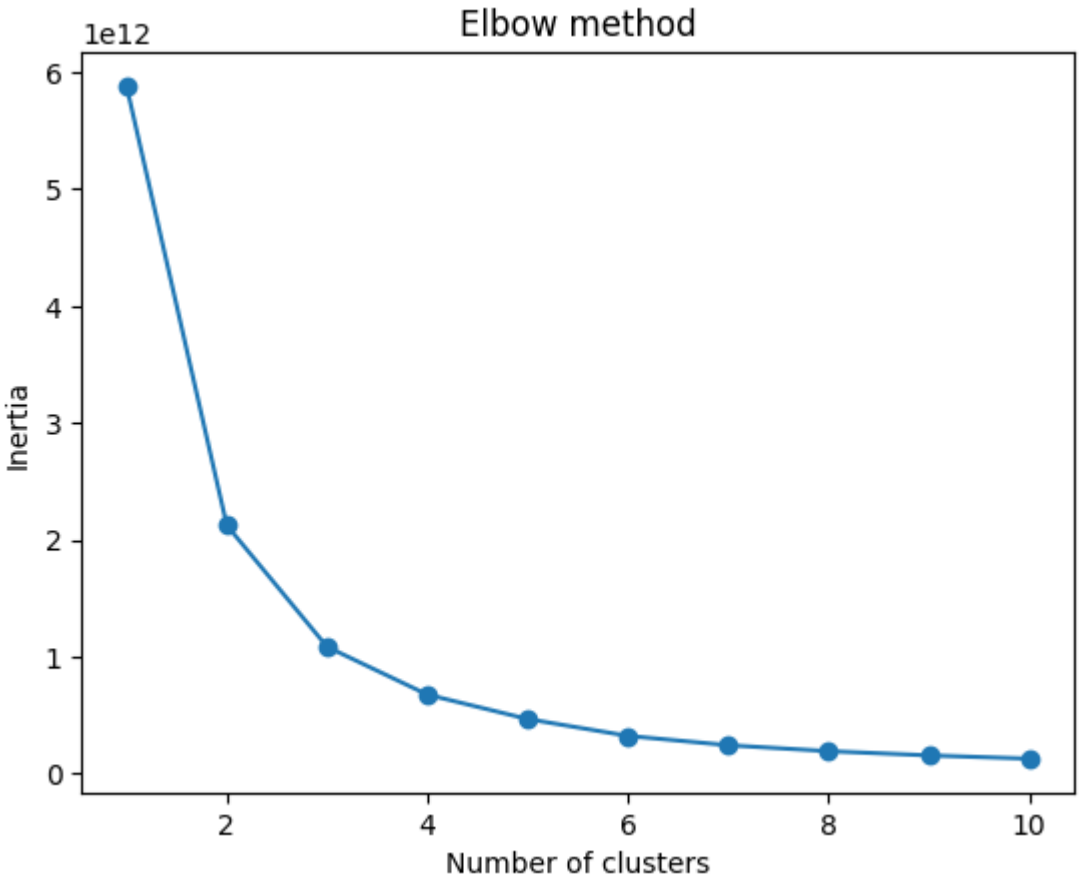
	model	year	price	transmission	mileage	fuelType	tax	mpg	engineSize
0	SLK	2005	5200	Automatic	63000	Petrol	325	32.1	1.8
1	S Class	2017	34948	Automatic	27000	Hybrid	20	61.4	2.1
2	SL CLASS	2016	49948	Automatic	6200	Petrol	555	28.0	5.5
3	G Class	2016	61948	Automatic	16000	Petrol	325	30.4	4.0
4	G Class	2016	73948	Automatic	4000	Petrol	325	30.1	4.0
...	...	...	...	...	...	...	...	...	...
13114	C Class	2020	35999	Automatic	500	Diesel	145	55.4	2.0
13115	B Class	2020	24699	Automatic	2500	Diesel	145	55.4	2.0
13116	GLC Class	2019	30999	Automatic	11612	Diesel	145	41.5	2.1
13117	CLS Class	2019	37990	Automatic	2426	Diesel	145	45.6	2.0
13118	S Class	2019	54999	Automatic	2075	Diesel	145	52.3	2.9

13119 rows × 9 columns

In [6]:

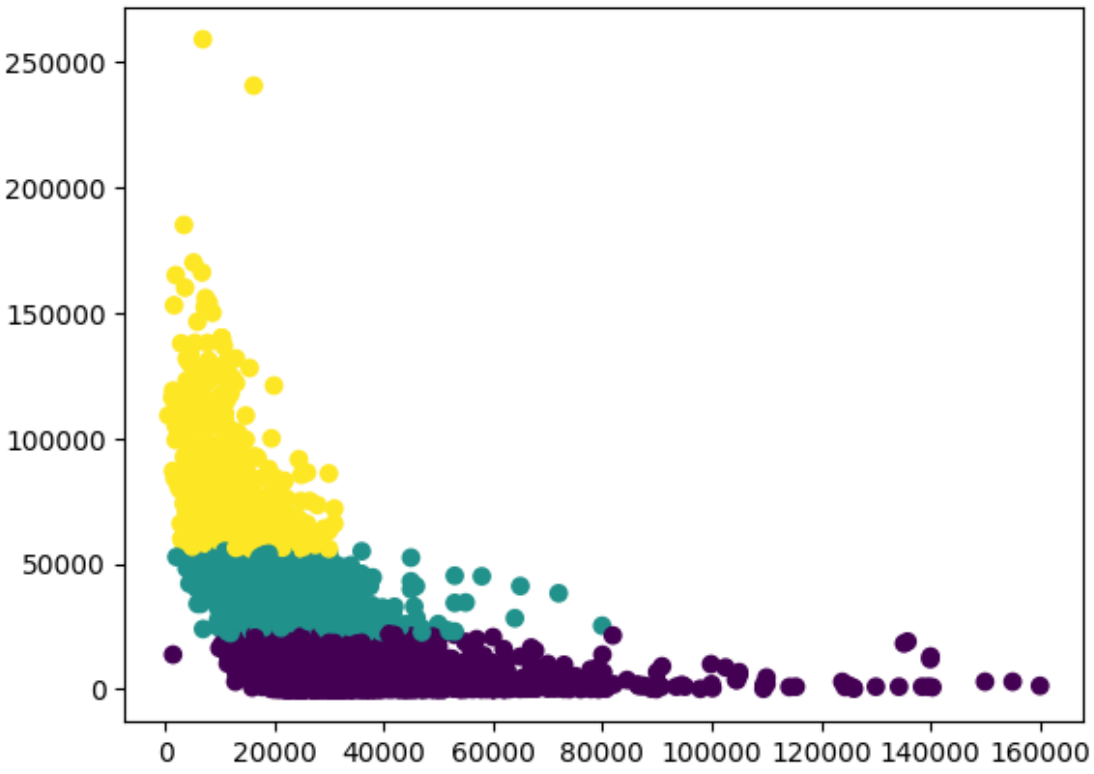
```
from sklearn.cluster import KMeans

data = list(zip(np.log(df_merc['price']), df_merc['mileage']))
inertias = []

for i in range(1,11):
    kmeans = KMeans(n_clusters=i)
    kmeans.fit(data)
    inertias.append(kmeans.inertia_)

plt.plot(range(1,11), inertias, marker='o')
plt.title('Elbow method')
plt.xlabel('Number of clusters')
plt.ylabel('Inertia')
plt.show()
```

```
C:\Users\user\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'a
uto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\user\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'a
uto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\user\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'a
uto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\user\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'a
uto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\user\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'a
uto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\user\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'a
uto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\user\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'a
uto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\user\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'a
uto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\user\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'a
uto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
C:\Users\user\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'a
uto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
```


Elbow method
1e12
6
5
4
3
2
1
0
Inertia
2
4
6
8
10
Number of clusters

In [31]:
```
kmeans = KMeans(n_clusters=3)
kmeans.fit(data)

plt.scatter(df_merc['price'], df_merc['mileage'], c=kmeans.labels_)
plt.show()
```

```
C:\Users\user\Anaconda3\lib\site-packages\sklearn\cluster\_kmeans.py:141
2: FutureWarning: The default value of `n_init` will change from 10 to 'a
uto' in 1.4. Set the value of `n_init` explicitly to suppress the warning
  super()._check_params_vs_input(X, default_n_init=10)
```



In [ ]:

In [ ]:

In [15]:
```
df1 = pd.read_csv('C:/Users/user/Desktop/My learning/ClinSoft/healthcare-d
df1
```

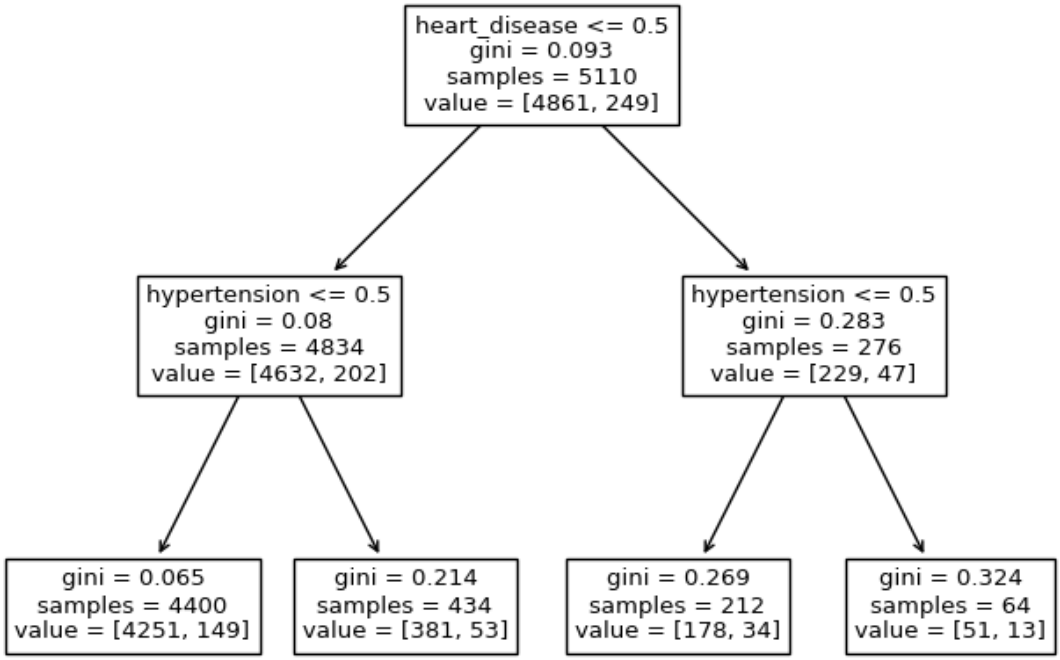
Out[15]:

	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residenc
0	9046	Male	67.0	0	1	Yes	Private	
1	51676	Female	61.0	0	0	Yes	Self-employed	
2	31112	Male	80.0	0	1	Yes	Private	
3	60182	Female	49.0	0	0	Yes	Private	
4	1665	Female	79.0	1	0	Yes	Self-employed	
...	...	...	...	...	...	...	...	
5105	18234	Female	80.0	1	0	Yes	Private	
5106	44873	Female	81.0	0	0	Yes	Self-employed	
5107	19723	Female	35.0	0	0	Yes	Self-employed	
5108	37544	Male	51.0	0	0	Yes	Private	
5109	44679	Female	44.0	0	0	Yes	Govt_job	

5110 rows × 12 columns

In [27]:
```
import pandas
from sklearn import tree
from sklearn.tree import DecisionTreeClassifier
import matplotlib.pyplot as plt



fig, ax = plt.subplots(figsize=(8, 6))
features = ['hypertension', 'heart_disease', ]
X = df1[features]
y = df1['stroke']

dtree = DecisionTreeClassifier()
dtree = dtree.fit(X, y)

tree.plot_tree(dtree, feature_names=features)
```

Out[27]:
```
[Text(0.5, 0.8333333333333334, 'heart_disease <= 0.5\ngini = 0.093\nsampl
es = 5110\nvalue = [4861, 249]'),
 Text(0.25, 0.5, 'hypertension <= 0.5\ngini = 0.08\nsamples = 4834\nvalue
= [4632, 202]'),
 Text(0.125, 0.16666666666666666, 'gini = 0.065\nsamples = 4400\nvalue =
[4251, 149]'),
 Text(0.375, 0.16666666666666666, 'gini = 0.214\nsamples = 434\nvalue =
[381, 53]'),
 Text(0.75, 0.5, 'hypertension <= 0.5\ngini = 0.283\nsamples = 276\nvalue
= [229, 47]'),
 Text(0.625, 0.16666666666666666, 'gini = 0.269\nsamples = 212\nvalue =
[178, 34]'),
 Text(0.875, 0.16666666666666666, 'gini = 0.324\nsamples = 64\nvalue = [5
1, 13]')]
```


heart_disease <= 0.5
gini = 0.093
samples = 5110
value = [4861, 249]
hypertension <= 0.5
gini = 0.08
samples = 4834
value = [4632, 202]
hypertension <= 0.5
gini = 0.283
samples = 276
value = [229, 47]
gini = 0.065
samples = 4400
value = [4251, 149]
gini = 0.214
samples = 434
value = [381, 53]
gini = 0.269
samples = 212
value = [178, 34]
gini = 0.324
samples = 64
value = [51, 13]