In [24]:

```
import pandas as pd
import numpy as np
```

In [3]:

```
df = pd.read_csv('cars - cars.csv')
df.head()
```

Out[3]:

	brand	km_driven	fuel	owner	selling_price
0	Maruti	145500	Diesel	First Owner	450000
1	Skoda	120000	Diesel	Second Owner	370000
2	Honda	140000	Petrol	Third Owner	158000
3	Hyundai	127000	Diesel	First Owner	225000
4	Maruti	120000	Petrol	First Owner	130000

In [5]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8128 entries, 0 to 8127
Data columns (total 5 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   brand          8128 non-null   object
 1   km_driven      8128 non-null   int64 
 2   fuel           8128 non-null   object
 3   owner          8128 non-null   object
 4   selling_price  8128 non-null   int64 
dtypes: int64(2), object(3)
memory usage: 317.6+ KB
```

In [7]:

```
df['fuel'].unique()
```

Out[7]:

```
array(['Diesel', 'Petrol', 'LPG', 'CNG'], dtype=object)
```

In [6]:

```
pd.get_dummies(df,columns=['owner','fuel']) #categorical to numeric
```

Out[6]:

	brand	km_driven	selling_price	owner_First Owner	owner_Fourth & Above Owner	owner_Second Owner	owner_Te Drive C
0	Maruti	145500	450000	1	0	0	
1	Skoda	120000	370000	0	0	1	
2	Honda	140000	158000	0	0	0	
3	Hyundai	127000	225000	1	0	0	
4	Maruti	120000	130000	1	0	0	
...	...	...	...	...	...	...	
8123	Hyundai	110000	320000	1	0	0	
8124	Hyundai	119000	135000	0	1	0	
8125	Maruti	120000	382000	1	0	0	
8126	Tata	25000	290000	1	0	0	
8127	Tata	25000	290000	1	0	0	

8128 rows × 12 columns

In [21]:

```
df.shape
```

Out[21]:

```
(8128, 5)
```

In [8]:

```
#dummy variable trap
```

In [9]:

```
from sklearn.preprocessing import OneHotEncoder
```

In [11]:

```
ohe  =OneHotEncoder(drop='first',sparse=False,dtype=np.int32)
```

In [13]:

```
x_train_new = ohe.fit_transform(df[['fuel','owner']])
```

In [14]:

```
x_train_new
```

Out[14]:

```
array([[1, 0, 0, ..., 0, 0, 0],
       [1, 0, 0, ..., 1, 0, 0],
       [0, 0, 1, ..., 0, 0, 1],
       ...,
       [1, 0, 0, ..., 0, 0, 0],
       [1, 0, 0, ..., 0, 0, 0],
       [1, 0, 0, ..., 0, 0, 0]])
```

In [15]:

```
from sklearn.preprocessing import LabelEncoder
```

In [16]:

```
le = LabelEncoder()
```

In [19]:

```
le_new = le.fit_transform(df['fuel'])
```

In [20]:

```
pd.DataFrame(le_new)
```

Out[20]:

	0
0	1
1	1
2	3
3	1
4	3
...	...
8123	3
8124	1
8125	1
8126	1
8127	1

8128 rows × 1 columns

In [ ]:

```

```