In [1]:
```
import pandas as pd
import numpy as np
```

In [11]:
```
''' Dataset info :
 Old Car Price Data (https://www.kaggle.com/datasets/rickyrrii/old-car-pr
'''
```

Out[11]:
```
' Dataset info : \n Old Car Price Data (https://www.kaggle.com/datasets/r
ickyrrii/old-car-price-data)\n'
```

In [1]:
```
df = pd.read_csv('imports-85.data.txt')
```

In [2]:
```
df
```

Out[2]:

	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.60	...	130	mpfi	3.47
0	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47
1	1	?	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68
2	2	164	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19
3	2	164	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19
4	2	?	audi	gas	std	two	sedan	fwd	front	99.8	...	136	mpfi	3.19
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
199	-1	95	volvo	gas	std	four	sedan	rwd	front	109.1	...	141	mpfi	3.78
200	-1	95	volvo	gas	turbo	four	sedan	rwd	front	109.1	...	141	mpfi	3.78
201	-1	95	volvo	gas	std	four	sedan	rwd	front	109.1	...	173	mpfi	3.58
202	-1	95	volvo	diesel	turbo	four	sedan	rwd	front	109.1	...	145	idi	3.01
203	-1	95	volvo	gas	turbo	four	sedan	rwd	front	109.1	...	141	mpfi	3.78

204 rows × 26 columns

In [3]:
```
df.head()
```

Out[3]:

	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.60	...	130	mpfi	3.47	2.68
0	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68
1	1	?	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68	3.47
2	2	164	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19	3.40
3	2	164	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19	3.40
4	2	?	audi	gas	std	two	sedan	fwd	front	99.8	...	136	mpfi	3.19	3.40

5 rows × 26 columns

In [4]:
```
df.tail()
```

Out[4]:

	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.60	...	130	mpfi	3.47	2
199	-1	95	volvo	gas	std	four	sedan	rwd	front	109.1	...	141	mpfi	3.78	3
200	-1	95	volvo	gas	turbo	four	sedan	rwd	front	109.1	...	141	mpfi	3.78	3
201	-1	95	volvo	gas	std	four	sedan	rwd	front	109.1	...	173	mpfi	3.58	2
202	-1	95	volvo	diesel	turbo	four	sedan	rwd	front	109.1	...	145	idi	3.01	3
203	-1	95	volvo	gas	turbo	four	sedan	rwd	front	109.1	...	141	mpfi	3.78	3

5 rows × 26 columns

In [5]:
```
df.dropna()
```

Out[5]:

	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.60	...	130	mpfi	3.47
0	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47
1	1	?	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68
2	2	164	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19
3	2	164	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19
4	2	?	audi	gas	std	two	sedan	fwd	front	99.8	...	136	mpfi	3.19
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
199	-1	95	volvo	gas	std	four	sedan	rwd	front	109.1	...	141	mpfi	3.78
200	-1	95	volvo	gas	turbo	four	sedan	rwd	front	109.1	...	141	mpfi	3.78
201	-1	95	volvo	gas	std	four	sedan	rwd	front	109.1	...	173	mpfi	3.58
202	-1	95	volvo	diesel	turbo	four	sedan	rwd	front	109.1	...	145	idi	3.01
203	-1	95	volvo	gas	turbo	four	sedan	rwd	front	109.1	...	141	mpfi	3.78

204 rows × 26 columns

In [6]:
```
df.describe()
```

Out[6]:

	3	88.60	168.80	64.10	48.80	2548	130
count	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000	204.000000
mean	0.823529	98.806373	174.075000	65.916667	53.749020	2555.602941	126.892157
std	1.239035	5.994144	12.362123	2.146716	2.424901	521.960820	41.744569
min	-2.000000	86.600000	141.100000	60.300000	47.800000	1488.000000	61.000000
25%	0.000000	94.500000	166.300000	64.075000	52.000000	2145.000000	97.000000
50%	1.000000	97.000000	173.200000	65.500000	54.100000	2414.000000	119.500000
75%	2.000000	102.400000	183.200000	66.900000	55.500000	2939.250000	142.000000
max	3.000000	120.900000	208.100000	72.300000	59.800000	4066.000000	326.000000

In [7]:
```
df.isnull()
```

Out[7]:

	3	?	alfa-romero	gas	std	two	convertible	rwd	front	88.60	...	130	mp
0	False	False	False	False	False	False	False	False	False	False	...	False	Fals
1	False	False	False	False	False	False	False	False	False	False	...	False	Fals
2	False	False	False	False	False	False	False	False	False	False	...	False	Fals
3	False	False	False	False	False	False	False	False	False	False	...	False	Fals
4	False	False	False	False	False	False	False	False	False	False	...	False	Fals
...	...	...	...	...	...	...	...	...	...	...	...	...	
199	False	False	False	False	False	False	False	False	False	False	...	False	Fals
200	False	False	False	False	False	False	False	False	False	False	...	False	Fals
201	False	False	False	False	False	False	False	False	False	False	...	False	Fals
202	False	False	False	False	False	False	False	False	False	False	...	False	Fals
203	False	False	False	False	False	False	False	False	False	False	...	False	Fals

204 rows × 26 columns

In [9]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 204 entries, 0 to 203
Data columns (total 26 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   3            204 non-null    int64  
 1   ?            204 non-null    object 
 2   alfa-romero  204 non-null    object 
 3   gas          204 non-null    object 
 4   std          204 non-null    object 
 5   two          204 non-null    object 
 6   convertible  204 non-null    object 
 7   rwd          204 non-null    object 
 8   front        204 non-null    object 
 9   88.60        204 non-null    float64
 10  168.80       204 non-null    float64
 11  64.10        204 non-null    float64
 12  48.80        204 non-null    float64
 13  2548         204 non-null    int64  
 14  dohc         204 non-null    object 
 15  four         204 non-null    object 
 16  130          204 non-null    int64  
 17  mpfi         204 non-null    object 
 18  3.47         204 non-null    object 
 19  2.68         204 non-null    object 
 20  9.00         204 non-null    float64
 21  111          204 non-null    object 
 22  5000         204 non-null    object 
 23  21           204 non-null    int64  
 24  27           204 non-null    int64  
 25  13495        204 non-null    object 
dtypes: float64(5), int64(5), object(16)
memory usage: 41.6+ KB
```

In [14]:
```
df = df.astype({'gas':'string'})
```

In [15]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 204 entries, 0 to 203
Data columns (total 26 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   3            204 non-null    int64  
 1   ?            204 non-null    object 
 2   alfa-romero  204 non-null    object 
 3   gas          204 non-null    string 
 4   std          204 non-null    object 
 5   two          204 non-null    object 
 6   convertible  204 non-null    object 
 7   rwd          204 non-null    object 
 8   front        204 non-null    object 
 9   88.60        204 non-null    float64
 10  168.80       204 non-null    float64
 11  64.10        204 non-null    float64
 12  48.80        204 non-null    float64
 13  2548         204 non-null    int64  
 14  dohc         204 non-null    object 
 15  four         204 non-null    object 
 16  130          204 non-null    int64  
 17  mpfi         204 non-null    object 
 18  3.47         204 non-null    object 
 19  2.68         204 non-null    object 
 20  9.00         204 non-null    float64
 21  111          204 non-null    object 
 22  5000         204 non-null    object 
 23  21           204 non-null    int64  
 24  27           204 non-null    int64  
 25  13495        204 non-null    object 
dtypes: float64(5), int64(5), object(15), string(1)
memory usage: 41.6+ KB
```

In [17]:
```
pd.get_dummies(df['gas'])
```

Out[17]:

	diesel	gas
0	0	1
1	0	1
2	0	1
3	0	1
4	0	1
...	...	...
199	0	1
200	0	1
201	0	1
202	1	0
203	0	1

204 rows × 2 columns

In [ ]: