

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

```
In [2]: df = pd.read_csv("D:\\IPSITA\\Automobile_data.csv")
```

```
In [3]: null_values = df.isnull().sum()
print("Null Values in the Dataset:")
print(null_values)
```

Null Values in the Dataset:

symboling	0
normalized-losses	0
make	0
fuel-type	0
aspiration	0
num-of-doors	0
body-style	0
drive-wheels	0
engine-location	0
wheel-base	0
length	0
width	0
height	0
curb-weight	0
engine-type	0
num-of-cylinders	0
engine-size	0
fuel-system	0
bore	0
stroke	0
compression-ratio	0
horsepower	0
peak-rpm	0
city-mpg	0
highway-mpg	0
price	0

dtype: int64

```
In [4]: count_question_mark = (df['normalized-losses'] == '?').sum()
```

```
In [5]: count_question_mark
```

Out[5]: 41

```
In [8]: df['normalized-losses'] = pd.to_numeric(df['normalized-losses'])
```

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

	symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	...	engine-size	fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg	highway-mpg	price
0	3	NaN	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111	5000	21	27	13495
1	3	NaN	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111	5000	21	27	16500
2	1	NaN	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68	3.47	9.0	154	5000	19	26	16500
3	2	164.0	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19	3.4	10.0	102	5500	24	30	13950
4	2	164.0	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19	3.4	8.0	115	5500	18	22	17450

5 rows × 26 columns

```
In [10]: df['normalized-losses'].fillna(df['normalized-losses'].mean(), inplace=True)
```

```
In [11]: df.head()
```

	symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	...	engine-size	fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg	highway-mpg	price
0	3	122.0	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111	5000	21	27	13495
1	3	122.0	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111	5000	21	27	16500
2	1	122.0	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68	3.47	9.0	154	5000	19	26	16500
3	2	164.0	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19	3.4	10.0	102	5500	24	30	13950
4	2	164.0	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19	3.4	8.0	115	5500	18	22	17450

5 rows × 26 columns

```
In [16]: def iqr(df, horsepower):
    Q1 = df[horsepower].quantile(0.25)
    Q3 = df[horsepower].quantile(0.75)
    IQR = Q3 - Q1

    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR

    outliers = df[(df[horsepower] < lower_bound) | (df[horsepower] > upper_bound)]
    cleaned_df = df[~((df[horsepower] < lower_bound) | (df[horsepower] > upper_bound))]

    return cleaned_df, outliers
```

```
In [23]: df.head(13)
```

	symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	...	engine-size	fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg	highway-mpg	price
0	3	122.0	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111	5000	21	27	13495
1	3	122.0	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111	5000	21	27	16500
2	1	122.0	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68	3.47	9.0	154	5000	19	26	16500
3	2	164.0	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19	3.4	10.0	102	5500	24	30	13950
4	2	164.0	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19	3.4	8.0	115	5500	18	22	17450
5	2	122.0	audi	gas	std	two	sedan	fwd	front	99.8	...	136	mpfi	3.19	3.4	8.5	110	5500	19	25	15250
6	1	158.0	audi	gas	std	four	sedan	fwd	front	105.8	...	136	mpfi	3.19	3.4	8.5	110	5500	19	25	17710
7	1	122.0	audi	gas	std	four	wagon	fwd	front	105.8	...	136	mpfi	3.19	3.4	8.5	110	5500	19	25	18920
8	1	158.0	audi	gas	turbo	four	sedan	fwd	front	105.8	...	131	mpfi	3.13	3.4	8.3	140	5500	17	20	23875
9	0	122.0	audi	gas	turbo	two	hatchback	4wd	front	99.5	...	131	mpfi	3.13	3.4	7.0	160	5500	16	22	?
10	2	192.0	bmw	gas	std	two	sedan	rwd	front	101.2	...	108	mpfi	3.5	2.8	8.8	101	5800	23	29	16430
11	0	192.0	bmw	gas	std	four	sedan	rwd	front	101.2	...	108	mpfi	3.5	2.8	8.8	101	5800	23	29	16925
12	0	188.0	bmw	gas	std	two	sedan	rwd	front	101.2	...	164	mpfi	3.31	3.19	9.0	121	4250	21	28	20970

13 rows × 26 columns

```
In [27]: df['horsepower'] = pd.to_numeric(df['horsepower'],errors='coerce')
cleaned_df, outliers = iqr(df, 'horsepower')
```

```
In [28]: cleaned_df
```

	symboling	normalized-losses	make	fuel-type	aspiration	num-of-doors	body-style	drive-wheels	engine-location	wheel-base	...	engine-size	fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	city-mpg	highway-mpg	price
0	3	122.0	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111.0	5000	21	27	13495
1	3	122.0	alfa-romero	gas	std	two	convertible	rwd	front	88.6	...	130	mpfi	3.47	2.68	9.0	111.0	5000	21	27	16500
2	1	122.0	alfa-romero	gas	std	two	hatchback	rwd	front	94.5	...	152	mpfi	2.68	3.47	9.0	154.0	5000	19	26	16500
3	2	164.0	audi	gas	std	four	sedan	fwd	front	99.8	...	109	mpfi	3.19	3.4	10.0	102.0	5500	24	30	13950
4	2	164.0	audi	gas	std	four	sedan	4wd	front	99.4	...	136	mpfi	3.19	3.4	8.0	115.0	5500	18	22	17450
...
200	-1	95.0	volvo	gas	std	four	sedan	rwd	front	109.1	...	141	mpfi	3.78	3.15	9.5	114.0	5400	23	28	16845
201	-1	95.0	volvo	gas	turbo	four	sedan	rwd	front	109.1	...	141	mpfi	3.78	3.15	8.7	160.0	5300	19	25	19045
202	-1	95.0	volvo	gas	std	four	sedan	rwd	front	109.1	...	173	mpfi	3.58	2.87	8.8	134.0	5500	18	23	21485
203	-1	95.0	volvo	diesel	turbo	four	sedan	rwd	front	109.1	...	145	idi	3.01	3.4	23.0	106.0	4800	26	27	22470
204	-1	95.0	volvo	gas	turbo	four	sedan	rwd	front	109.1	...	141	mpfi	3.78	3.15	9.5	114.0	5400	19	25	22625

199 rows × 26 columns

```
In [29]: cleaned_df.to_csv('automobile_data_new_file.csv')
```

```
In [ ]:
```