Cleaning the Dataset

▾ Cleaning the Dataset

```python
[7]  car=car[car['year'].str.isnumeric()]
```

```python
[8]  car['year']=car['year'].astype(int)
```

```python
[9]  car=car[car['Price']!='Ask For Price']
```

```python
[10] car['Price']=car['Price'].str.replace(',','').astype(int)
```

```python
[11] car['kms_driven']=car['kms_driven'].str.split().str.get(0).str.replace(',','')
```

```python
[12] car=car[car['kms_driven'].str.isnumeric()]
```

```python
[13] car['kms_driven']=car['kms_driven'].astype(int)
```

```python
[14] car=car[~car['fuel_type'].isna()]
```

```python
[15] car.shape
```
```
(816, 6)
```

```python
[16] car['name']=car['name'].str.split().str.slice(start=0,stop=3).str.join(' ')
```

```python
[17] car=car.reset_index(drop=True)
```

```
[18] car
```

	name	company	year	Price	kms_driven	fuel_type
0	Hyundai Santro Xing	Hyundai	2007	80000	45000	Petrol
1	Mahindra Jeep CL550	Mahindra	2006	425000	40	Diesel
2	Hyundai Grand i10	Hyundai	2014	325000	28000	Petrol
3	Ford EcoSport Titanium	Ford	2014	575000	36000	Diesel
4	Ford Figo	Ford	2012	175000	41000	Diesel
...	...	...	...	...	...	...
811	Maruti Suzuki Ritz	Maruti	2011	270000	50000	Petrol
812	Tata Indica V2	Tata	2009	110000	30000	Diesel
813	Toyota Corolla Altis	Toyota	2009	300000	132000	Petrol
814	Tata Zest XM	Tata	2018	260000	27000	Diesel
815	Mahindra Quanto C8	Mahindra	2013	390000	40000	Diesel

816 rows × 6 columns

```
car.to_csv('Cleaned_Car_data.csv')
```

```
[20] car.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 816 entries, 0 to 815
Data columns (total 6 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   name        816 non-null    object
 1   company     816 non-null    object
 2   year        816 non-null    int64
 3   Price       816 non-null    int64
 4   kms_driven  816 non-null    int64
 5   fuel_type   816 non-null    object
dtypes: int64(3), object(3)
memory usage: 38.4+ KB
```

```
[21] car.describe(include='all')
```

	name	company	year	Price	kms_driven	fuel_type
count	816	816	816.000000	8.160000e+02	816.000000	816
unique	254	25	NaN	NaN	NaN	3
top	Maruti Suzuki Swift	Maruti	NaN	NaN	NaN	Petrol
freq	51	221	NaN	NaN	NaN	428
mean	NaN	NaN	2012.444853	4.117176e+05	46275.531863	NaN
std	NaN	NaN	4.002992	4.751844e+05	34297.428044	NaN
min	NaN	NaN	1995.000000	3.000000e+04	0.000000	NaN
25%	NaN	NaN	2010.000000	1.750000e+05	27000.000000	NaN
50%	NaN	NaN	2013.000000	2.999990e+05	41000.000000	NaN
75%	NaN	NaN	2015.000000	4.912500e+05	56818.500000	NaN
max	NaN	NaN	2019.000000	8.500003e+06	400000.000000	NaN

```
[22] car=car[car['Price']<6000000]
```