

```
In [20]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [21]: data=pd.read_csv(r"C:\Users\akshi\OneDrive\Desktop\CAR DETAILS FROM CAR DEKH")
```

```
In [22]: from sklearn.preprocessing import LabelEncoder
```

```
In [23]: le=LabelEncoder()
```

```
In [24]: data.name=le.fit_transform(data.name)
```

```
In [25]: data.fuel=le.fit_transform(data.fuel)
```

```
In [26]: data.seller_type=le.fit_transform(data.seller_type)
```

```
In [27]: data.transmission=le.fit_transform(data.transmission)
```

```
In [28]: data.owner=le.fit_transform(data.owner)
```

```
In [29]: data.head()
```

```
Out[29]:
```

	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner
0	775	2007	60000	70000	4	1	1	0
1	1041	2007	135000	50000	4	1	1	0
2	505	2012	600000	100000	1	1	1	0
3	118	2017	250000	46000	4	1	1	0
4	279	2014	450000	141000	1	1	1	2

```
In [30]: data.tail()
```

```
Out[30]:
```

	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner
4335	602	2014	409999	80000	1	1	1	2
4336	601	2014	409999	80000	1	1	1	2
4337	777	2009	110000	83000	4	1	1	2

	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner
4338	381	2016	865000	90000	1	1	1	0
4339	1154	2016	225000	40000	4	1	1	0

In [31]: `data.describe()`

Out[31]:

	name	year	selling_price	km_driven	fuel	seller_type	t
count	4340.000000	4340.000000	4.340000e+03	4340.000000	4340.000000	4340.000000	
mean	751.591244	2013.090783	5.041273e+05	66215.777419	2.469124	0.794470	
std	400.046963	4.215344	5.785487e+05	46644.102194	1.508435	0.458629	
min	0.000000	1992.000000	2.000000e+04	1.000000	0.000000	0.000000	
25%	432.000000	2011.000000	2.087498e+05	35000.000000	1.000000	1.000000	
50%	793.000000	2014.000000	3.500000e+05	60000.000000	1.000000	1.000000	
75%	1048.000000	2016.000000	6.000000e+05	90000.000000	4.000000	1.000000	
max	1490.000000	2020.000000	8.900000e+06	806599.000000	4.000000	2.000000	

In [32]: `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4340 entries, 0 to 4339
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   name             4340 non-null   int32
1   year             4340 non-null   int64
2   selling_price    4340 non-null   int64
3   km_driven        4340 non-null   int64
4   fuel             4340 non-null   int32
5   seller_type      4340 non-null   int32
6   transmission     4340 non-null   int32
7   owner            4340 non-null   int32
dtypes: int32(5), int64(3)
memory usage: 186.6 KB
```

In [33]: `data.corr()`

Out[33]:

	name	year	selling_price	km_driven	fuel	seller_type	transmission
name	1.000000	-0.051902	-0.077598	0.126203	-0.083927	0.140802	0.08
year	-0.051902	1.000000	0.413922	-0.419688	-0.120002	-0.098352	-0.14
selling_price	-0.077598	0.413922	1.000000	-0.192289	-0.269653	-0.151554	-0.53
km_driven	0.126203	-0.419688	-0.192289	1.000000	-0.286095	0.113689	0.12

	name	year	selling_price	km_driven	fuel	seller_type	transmission
fuel	-0.083927	-0.120002	-0.269653	-0.286095	1.000000	0.038797	0.03
seller_type	0.140802	-0.098352	-0.151554	0.113689	0.038797	1.000000	0.17
transmission	0.087785	-0.143800	-0.530205	0.120226	0.039249	0.174925	1.00
owner	0.032819	-0.414705	-0.207840	0.297115	-0.010910	0.165681	0.07

In [34]:

```
data.columns
```

Out[34]:

```
Index(['name', 'year', 'selling_price', 'km_driven', 'fuel', 'seller_type',
      'transmission', 'owner'],
      dtype='object')
```

In [35]:

```
from os import X_OK
```

In [39]:

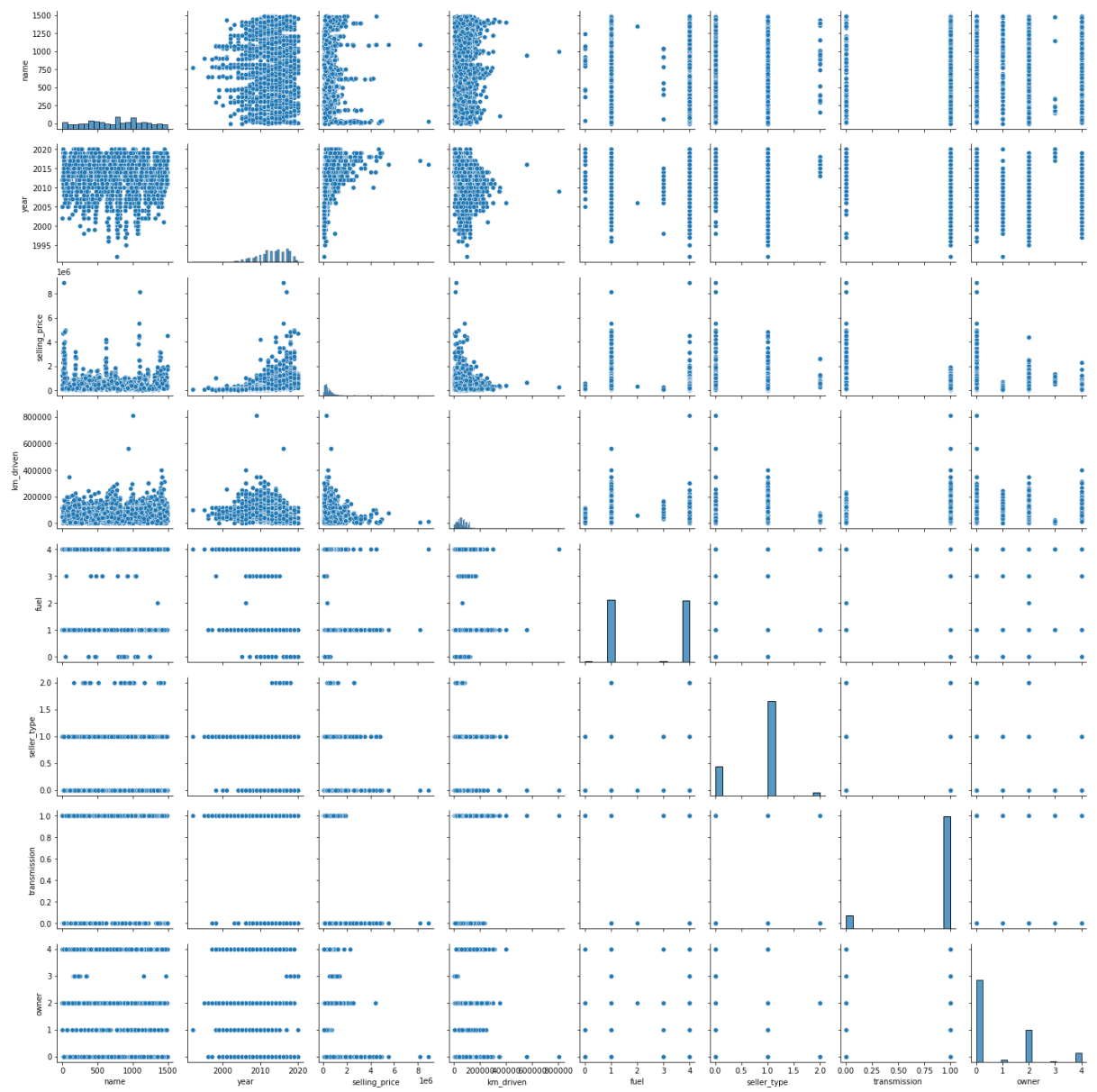
```
y=data['selling_price']
x=data[['name', 'year', 'km_driven', 'fuel', 'seller_type',
      'transmission', 'owner']]
```

In [40]:

```
sns.pairplot(data)
```

Out[40]:

```
<seaborn.axisgrid.PairGrid at 0x2129d3be790>
```



```
In [41]: from sklearn.linear_model import LinearRegression
```

```
In [43]: from sklearn.model_selection import train_test_split
```

```
In [44]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,train_size=
```

```
In [45]: x_train.head()
```

```
Out[45]:
```

	name	year	km_driven	fuel	seller_type	transmission	owner
2419	1377	2012	56000	1	1	1	0
300	551	2009	50000	4	1	1	0
34	1334	2018	15000	4	1	1	0
2733	177	2018	10000	4	1	0	0
3443	63	2011	90000	1	1	1	2

```
In [46]: x_test.head()
```

```
Out[46]:
```

	name	year	km_driven	fuel	seller_type	transmission	owner
63	1048	2013	80000	4	0	1	0
4	279	2014	141000	1	1	1	2
4133	543	2015	55000	1	0	1	0
3924	1370	2014	140000	1	1	1	2
577	434	2017	19890	4	0	0	0

```
In [47]: y_train.head()
```

```
Out[47]: 2419    409999
300      250000
34       525000
2733     875000
3443     335000
Name: selling_price, dtype: int64
```

```
In [48]: y_test.head()
```

```
Out[48]: 63      300000
4       450000
4133    399000
3924    500000
577     550000
Name: selling_price, dtype: int64
```

```
In [49]: model=LinearRegression()
```

```
In [50]: model.fit(x,y)
```

```
Out[50]: LinearRegression()
```

```
In [52]: model.intercept_
```

Out[52]: -72110142.97883077

In [53]: `model.coef_`

Out[53]: `array([-4.37398806e+01, 3.66420639e+04, -9.64602163e-01, -9.35713443e+04,
 -1.94361948e+04, -8.83466747e+05, -1.71082821e+04])`

In []: