

```
In [2]: import pandas as pd
```

```
In [3]: df=pd.read_csv('C:\\Users\\ADMIN\\Desktop\\Data sets For DS Projects\\Car-details-v3 (1).csv')
```

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

Out[4]:

	name	year	selling_price	km_driven	fuel	seller_type	transmission	owner	mileage	engine	max_power	torque	seats
0	Maruti Swift Dzire VDI	2014	450000	145500	Diesel	Individual	Manual	First Owner	23.4 kmpl	1248 CC	74 bhp	190Nm@ 2000rpm	5.0
1	Skoda Rapid 1.5 TDI Ambition	2014	370000	120000	Diesel	Individual	Manual	Second Owner	21.14 kmpl	1498 CC	103.52 bhp	250Nm@ 1500-2500rpm	5.0
2	Honda City 2017-2020 EXi	2006	158000	140000	Petrol	Individual	Manual	Third Owner	17.7 kmpl	1497 CC	78 bhp	12.7@ 2,700(kgm@ rpm)	5.0
3	Hyundai i20 Sportz Diesel	2010	225000	127000	Diesel	Individual	Manual	First Owner	23.0 kmpl	1396 CC	90 bhp	22.4 kgm at 1750-2750rpm	5.0
4	Maruti Swift VXi BSIII	2007	130000	120000	Petrol	Individual	Manual	First Owner	16.1 kmpl	1298 CC	88.2 bhp	11.5@ 4,500(kgm@ rpm)	5.0

```
In [5]: df.shape
```

Out[5]: (8128, 13)

```
In [10]: print(df['seller_type'].unique())
print(df['transmission'].unique())
print(df['owner'].unique())
print(df['fuel'].unique())

['Individual' 'Dealer' 'Trustmark Dealer']
['Manual' 'Automatic']
['First Owner' 'Second Owner' 'Third Owner' 'Fourth & Above Owner'
 'Test Drive Car']
['Diesel' 'Petrol' 'LPG' 'CNG']
```

```
In [11]: df.isnull().sum()
```

Out[11]:

name	0
year	0
selling_price	0
km_driven	0
fuel	0
seller_type	0
transmission	0
owner	0
mileage	221
engine	221
max_power	215
torque	222
seats	221

dtype: int64

```
In [12]: df.describe()
```

Out[12]:

	year	selling_price	km_driven	seats
count	8128.000000	8.128000e+03	8.128000e+03	7907.000000
mean	2013.804011	6.382718e+05	6.981951e+04	5.416719
std	4.044249	8.062534e+05	5.655055e+04	0.959588
min	1983.000000	2.999900e+04	1.000000e+00	2.000000
25%	2011.000000	2.549990e+05	3.500000e+04	5.000000
50%	2015.000000	4.500000e+05	6.000000e+04	5.000000

75%	2017.000000	6.750000e+05	9.800000e+04	5.000000
max	2020.000000	1.000000e+07	2.360457e+06	14.000000

```
In [13]: df.columns
```

```
Out[13]: Index(['name', 'year', 'selling_price', 'km_driven', 'fuel', 'seller_type',
        'transmission', 'owner', 'mileage', 'engine', 'max_power', 'torque',
        'seats'],
        dtype='object')
```

```
In [19]: final_dataset=df[['year', 'selling_price', 'km_driven',
        'fuel', 'seller_type', 'transmission', 'owner']]
```

```
In [20]: final_dataset.head()
```

```
Out[20]:
```

	year	selling_price	km_driven	fuel	seller_type	transmission	owner
0	2014	450000	145500	Diesel	Individual	Manual	First Owner
1	2014	370000	120000	Diesel	Individual	Manual	Second Owner
2	2006	158000	140000	Petrol	Individual	Manual	Third Owner
3	2010	225000	127000	Diesel	Individual	Manual	First Owner
4	2007	130000	120000	Petrol	Individual	Manual	First Owner

```
In [21]: final_dataset['Current_Year']=2021
```

C:\Users\ADMIN\AppData\Local\Temp\ipykernel_27900\147258130.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
final_dataset['Current_Year']=2021
```

```
In [22]: final_dataset.head()
```

```
Out[22]:
```

	year	selling_price	km_driven	fuel	seller_type	transmission	owner	Current_Year
0	2014	450000	145500	Diesel	Individual	Manual	First Owner	2021
1	2014	370000	120000	Diesel	Individual	Manual	Second Owner	2021
2	2006	158000	140000	Petrol	Individual	Manual	Third Owner	2021
3	2010	225000	127000	Diesel	Individual	Manual	First Owner	2021
4	2007	130000	120000	Petrol	Individual	Manual	First Owner	2021

```
In [25]: final_dataset['no_year']=final_dataset['year']-final_dataset['year']
```

C:\Users\ADMIN\AppData\Local\Temp\ipykernel_27900\297843625.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
final_dataset['no_year']=final_dataset['year']-final_dataset['year']
```

```
In [26]: final_dataset.head()
```

```
Out[26]:
```

	year	selling_price	km_driven	fuel	seller_type	transmission	owner	Current_Year	no_year
0	2014	450000	145500	Diesel	Individual	Manual	First Owner	2021	0
1	2014	370000	120000	Diesel	Individual	Manual	Second Owner	2021	0

2	2006	158000	140000	Petrol	Individual	Manual	Third Owner	2021	0
3	2010	225000	127000	Diesel	Individual	Manual	First Owner	2021	0
4	2007	130000	120000	Petrol	Individual	Manual	First Owner	2021	0

In [28]: `final_dataset.drop(['year'],axis=1)`

Out[28]:

	selling_price	km_driven	fuel	seller_type	transmission	owner	Current_Year	no_year
0	450000	145500	Diesel	Individual	Manual	First Owner	2021	0
1	370000	120000	Diesel	Individual	Manual	Second Owner	2021	0
2	158000	140000	Petrol	Individual	Manual	Third Owner	2021	0
3	225000	127000	Diesel	Individual	Manual	First Owner	2021	0
4	130000	120000	Petrol	Individual	Manual	First Owner	2021	0
...
8123	320000	110000	Petrol	Individual	Manual	First Owner	2021	0
8124	135000	119000	Diesel	Individual	Manual	Fourth & Above Owner	2021	0
8125	382000	120000	Diesel	Individual	Manual	First Owner	2021	0
8126	290000	25000	Diesel	Individual	Manual	First Owner	2021	0
8127	290000	25000	Diesel	Individual	Manual	First Owner	2021	0

8128 rows × 8 columns

In [29]: `final_dataset.drop(['year'],axis=1,inplace=True)`

C:\Users\ADMIN\anaconda3\lib\site-packages\pandas\core\frame.py:4906: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
return super().drop(

In [30]: `final_dataset.head()`

Out[30]:

	selling_price	km_driven	fuel	seller_type	transmission	owner	Current_Year	no_year
0	450000	145500	Diesel	Individual	Manual	First Owner	2021	0
1	370000	120000	Diesel	Individual	Manual	Second Owner	2021	0
2	158000	140000	Petrol	Individual	Manual	Third Owner	2021	0
3	225000	127000	Diesel	Individual	Manual	First Owner	2021	0
4	130000	120000	Petrol	Individual	Manual	First Owner	2021	0

In [32]: `final_dataset.drop(['no_year'],axis=1,inplace=True)`

C:\Users\ADMIN\anaconda3\lib\site-packages\pandas\core\frame.py:4906: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
return super().drop(

In [33]: `final_dataset.head()`

Out[33]:

	selling_price	km_driven	fuel	seller_type	transmission	owner	Current_Year
0	450000	145500	Diesel	Individual	Manual	First Owner	2021
1	370000	120000	Diesel	Individual	Manual	Second Owner	2021
2	158000	140000	Petrol	Individual	Manual	Third Owner	2021
3	225000	127000	Diesel	Individual	Manual	First Owner	2021

```
In [34]: final_dataset=pd.get_dummies(final_dataset,drop_first=True)
```

```
In [35]: final_dataset.head()
```

```
Out[35]:
```

	selling_price	km_driven	Current_Year	fuel_Diesel	fuel_LPG	fuel_Petrol	seller_type_Individual	seller_type_Trustmark Dealer	transmission_Manual
0	450000	145500	2021	1	0	0	1	0	1
1	370000	120000	2021	1	0	0	1	0	1
2	158000	140000	2021	0	0	1	1	0	1
3	225000	127000	2021	1	0	0	1	0	1
4	130000	120000	2021	0	0	1	1	0	1

```
In [36]: final_dataset.corr()
```

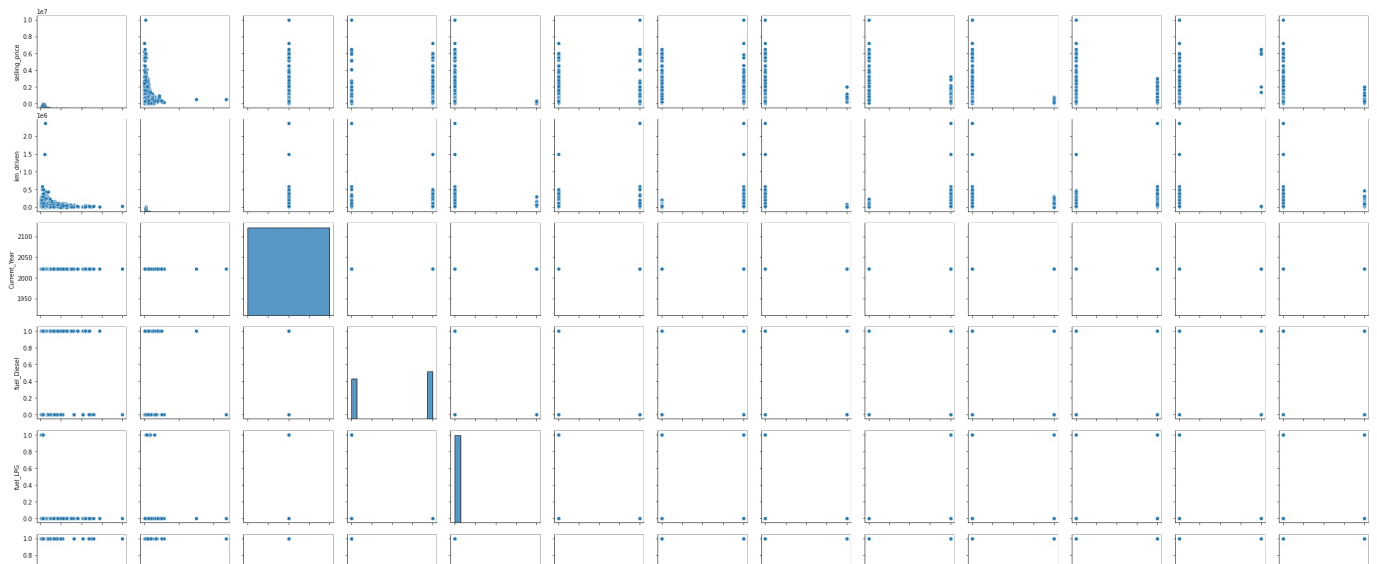
```
Out[36]:
```

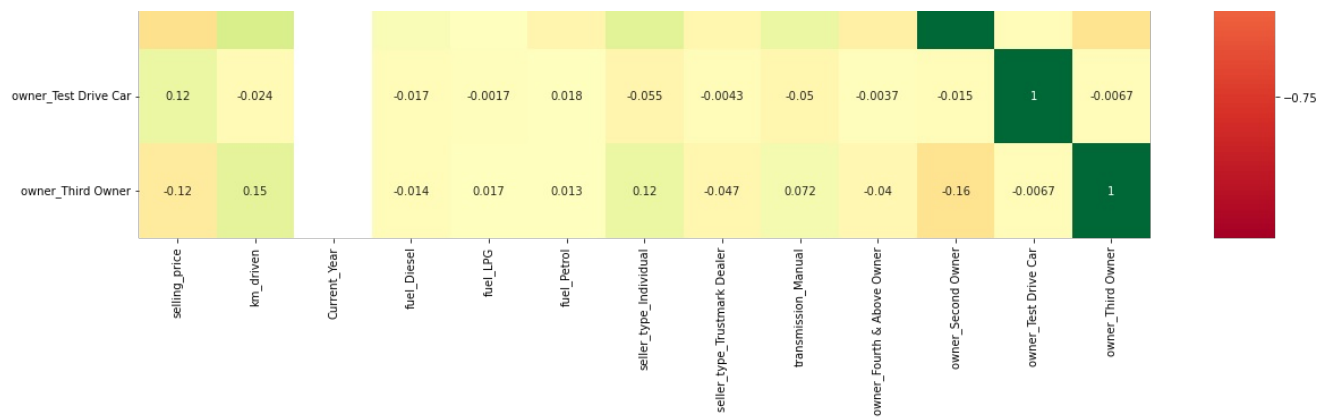
	selling_price	km_driven	Current_Year	fuel_Diesel	fuel_LPG	fuel_Petrol	seller_type_Individual	seller_type_Trustmark Dealer	tr
selling_price	1.000000	-0.225534	NaN	0.206521	-0.037222	-0.195975	-0.386244	0.035084	
km_driven	-0.225534	1.000000	NaN	0.266368	0.023458	-0.269802	0.204736	-0.084282	
Current_Year	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
fuel_Diesel	0.206521	0.266368	NaN	1.000000	-0.074494	-0.976687	-0.000239	-0.137961	
fuel_LPG	-0.037222	0.023458	NaN	-0.074494	1.000000	-0.061584	0.021092	-0.011852	
fuel_Petrol	-0.195975	-0.269802	NaN	-0.976687	-0.061584	1.000000	-0.008983	0.142335	
seller_type_Individual	-0.386244	0.204736	NaN	-0.000239	0.021092	-0.008983	1.000000	-0.385425	
seller_type_Trustmark Dealer	0.035084	-0.084282	NaN	-0.137961	-0.011852	0.142335	-0.385425	1.000000	
transmission_Manual	-0.588845	0.201229	NaN	0.025519	0.026397	-0.034631	0.375234	-0.149677	
owner_Fourth & Above Owner	-0.075669	0.094939	NaN	-0.010638	0.014785	0.007300	0.066360	-0.025577	
owner_Second Owner	-0.179881	0.209532	NaN	0.046762	0.017122	-0.051046	0.165983	-0.047034	
owner_Test Drive Car	0.115880	-0.024214	NaN	-0.017008	-0.001700	0.017630	-0.055297	-0.004290	
owner_Third Owner	-0.118956	0.145634	NaN	-0.014272	0.017199	0.012820	0.116236	-0.046814	

```
In [37]: import seaborn as sns
```

```
In [38]: sns.pairplot(final_dataset)
```

```
Out[38]: <seaborn.axisgrid.PairGrid at 0x1e1914ce100>
```





In [41]: `final_dataset.head()`

Out[41]:

	selling_price	km_driven	Current_Year	fuel_Diesel	fuel_LPG	fuel_Petrol	seller_type_Individual	seller_type_Trustmark Dealer	transmission_Manual
0	450000	145500	2021	1	0	0		1	0
1	370000	120000	2021	1	0	0		1	0
2	158000	140000	2021	0	0	1		1	0
3	225000	127000	2021	1	0	0		1	0
4	130000	120000	2021	0	0	1		1	0

In [43]: `x=final_dataset.iloc[:,1:]`
`y=final_dataset.iloc[:,0]`

In [44]: `x.head()`

Out[44]:

	km_driven	Current_Year	fuel_Diesel	fuel_LPG	fuel_Petrol	seller_type_Individual	seller_type_Trustmark Dealer	transmission_Manual	owner_Fourth & Above Owner
0	145500	2021	1	0	0		1	0	1
1	120000	2021	1	0	0		1	0	1
2	140000	2021	0	0	1		1	0	1
3	127000	2021	1	0	0		1	0	1
4	120000	2021	0	0	1		1	0	1

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

Out[45]:

```
0    450000
1    370000
2    158000
3    225000
4    130000
Name: selling_price, dtype: int64
```

In []: