In [5]:

**import** pandas **as** pd

In [6]:

df**=**pd**.**read\_csv('car data.csv')

In [7]:

df**.**head()

# Out[7]: Car\_Name Year Selling\_Price Present\_Price Kms\_Driven Fuel\_Type Seller\_Type Transmission Owner

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | ritz 2014 | 3.35 | 5.59 | 27000 | Petrol | Dealer | Manual | 0 |
| **1** | sx4 2013 | 4.75 | 9.54 | 43000 | Diesel | Dealer | Manual | 0 |
| **2** | ciaz 2017 | 7.25 | 9.85 | 6900 | Petrol | Dealer | Manual | 0 |
| **3** | wagon r 2011 | 2.85 | 4.15 | 5200 | Petrol | Dealer | Manual | 0 |
| **4** | swift 2014 | 4.60 | 6.87 | 42450 | Diesel | Dealer | Manual | 0 |

In [8]:

df**.**shape

Out[8]: (301, 9)

In [9]:

print(df['Seller\_Type']**.**unique()) print(df['Transmission']**.**unique()) print(df['Owner']**.**unique())

['Dealer' 'Individual'] ['Manual' 'Automatic'] [0 1 3]

In [10]:

df**.**isnull()**.**sum()

Out[10]: Car\_Name 0

Year 0

Selling\_Price 0

Present\_Price 0

Kms\_Driven 0

Fuel\_Type 0

Seller\_Type 0

Transmission 0

Owner 0

dtype: int64

In [11]:

df**.**columns

Out[11]: Index(['Car\_Name', 'Year', 'Selling\_Price', 'Present\_Price', 'Kms\_Driven', 'Fuel\_Type', 'Seller\_Type', 'Transmission', 'Owner'],

dtype='object')

In [12]:

final\_dataset**=**df[['Car\_Name', 'Year', 'Selling\_Price', 'Present\_Price', 'Kms\_Driven', 'Fuel\_Type', 'Seller\_Type', 'Transmission', 'Owner']]

In [13]:

final\_dataset**.**head()

# Out[13]: Car\_Name Year Selling\_Price Present\_Price Kms\_Driven Fuel\_Type Seller\_Type Transmission Owner

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | ritz 2014 | 3.35 | 5.59 | 27000 | Petrol | Dealer | Manual | 0 |
| **1** | sx4 2013 | 4.75 | 9.54 | 43000 | Diesel | Dealer | Manual | 0 |
| **2** | ciaz 2017 | 7.25 | 9.85 | 6900 | Petrol | Dealer | Manual | 0 |
| **3** | wagon r 2011 | 2.85 | 4.15 | 5200 | Petrol | Dealer | Manual | 0 |
| **4** | swift 2014 | 4.60 | 6.87 | 42450 | Diesel | Dealer | Manual | 0 |

In [14]:

final\_dataset**=**df[['Year', 'Selling\_Price', 'Present\_Price', 'Kms\_Driven', 'Fuel\_Type', 'Seller\_Type', 'Transmission', 'Owner']]

In [15]:

final\_dataset**.**head()

# Out[15]: Year Selling\_Price Present\_Price Kms\_Driven Fuel\_Type Seller\_Type Transmission Owner

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **0** 2014 | 3.35 | 5.59 | 27000 | Petrol | Dealer | Manual | 0 |
| **1** 2013 | 4.75 | 9.54 | 43000 | Diesel | Dealer | Manual | 0 |
| **2** 2017 | 7.25 | 9.85 | 6900 | Petrol | Dealer | Manual | 0 |
| **3** 2011 | 2.85 | 4.15 | 5200 | Petrol | Dealer | Manual | 0 |
| **4** 2014 | 4.60 | 6.87 | 42450 | Diesel | Dealer | Manual | 0 |

In [16]:

final\_dataset['Current\_Year']**=**2022

In [17]:

print(df['Seller\_Type']**.**unique())

['Dealer' 'Individual']

In [18]:

final\_dataset**.**head()

# Out[18]: Year Selling\_Price Present\_Price Kms\_Driven Fuel\_Type Seller\_Type Transmission Owner Current\_Year

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** 2014 | 3.35 | 5.59 | 27000 | Petrol | Dealer | Manual | 0 | 2022 |
| **1** 2013 | 4.75 | 9.54 | 43000 | Diesel | Dealer | Manual | 0 | 2022 |
| **2** 2017 | 7.25 | 9.85 | 6900 | Petrol | Dealer | Manual | 0 | 2022 |
| **3** 2011 | 2.85 | 4.15 | 5200 | Petrol | Dealer | Manual | 0 | 2022 |
| **4** 2014 | 4.60 | 6.87 | 42450 | Diesel | Dealer | Manual | 0 | 2022 |

In [19]:

final\_dataset['No.Year']**=**final\_dataset['Current\_Year']**-**final\_dataset['Year']

In [20]:

final\_dataset**.**head()

# Out[20]: Year Selling\_Price Present\_Price Kms\_Driven Fuel\_Type Seller\_Type Transmission Owner Current\_Year No.Year

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** 2014 | 3.35 | 5.59 | 27000 | Petrol | Dealer | Manual | 0 | 2022 | 8 |
| **1** 2013 | 4.75 | 9.54 | 43000 | Diesel | Dealer | Manual | 0 | 2022 | 9 |
| **2** 2017 | 7.25 | 9.85 | 6900 | Petrol | Dealer | Manual | 0 | 2022 | 5 |
| **3** 2011 | 2.85 | 4.15 | 5200 | Petrol | Dealer | Manual | 0 | 2022 | 11 |
| **4** 2014 | 4.60 | 6.87 | 42450 | Diesel | Dealer | Manual | 0 | 2022 | 8 |

In [21]:

final\_dataset**.**drop(['Year'],axis**=**1,inplace**=True**)

In [22]:

final\_dataset**.**head()

# Out[22]: Selling\_Price Present\_Price Kms\_Driven Fuel\_Type Seller\_Type Transmission Owner Current\_Year No.Year

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 3.35 | 5.59 | 27000 | Petrol | Dealer | Manual | 0 | 2022 | 8 |
| **1** | 4.75 | 9.54 | 43000 | Diesel | Dealer | Manual | 0 | 2022 | 9 |
| **2** | 7.25 | 9.85 | 6900 | Petrol | Dealer | Manual | 0 | 2022 | 5 |
| **3** | 2.85 | 4.15 | 5200 | Petrol | Dealer | Manual | 0 | 2022 | 11 |
| **4** | 4.60 | 6.87 | 42450 | Diesel | Dealer | Manual | 0 | 2022 | 8 |

In [23]:

final\_dataset**.**drop(['Current\_Year'],axis**=**1,inplace**=True**)

In [24]:

final\_dataset**.**head()

# Out[24]: Selling\_Price Present\_Price Kms\_Driven Fuel\_Type Seller\_Type Transmission Owner No.Year

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 3.35 | 5.59 | 27000 | Petrol | Dealer | Manual | 0 | 8 |
| **1** | 4.75 | 9.54 | 43000 | Diesel | Dealer | Manual | 0 | 9 |
| **2** | 7.25 | 9.85 | 6900 | Petrol | Dealer | Manual | 0 | 5 |
| **3** | 2.85 | 4.15 | 5200 | Petrol | Dealer | Manual | 0 | 11 |
| **4** | 4.60 | 6.87 | 42450 | Diesel | Dealer | Manual | 0 | 8 |

In [25]:

final\_dataset**=**pd**.**get\_dummies(final\_dataset,drop\_first**=True**)

In [26]:

final\_dataset**.**head()

# Out[26]: Selling\_Price Present\_Price Kms\_Driven Owner No.Year Fuel\_Type\_Diesel Fuel\_Type\_Petrol Seller\_Type\_Individual Transmission\_Manual

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 3.35 | 5.59 | 27000 | 0 | 8 | 0 | 1 | 0 | 1 |
| **1** | 4.75 | 9.54 | 43000 | 0 | 9 | 1 | 0 | 0 | 1 |
| **2** | 7.25 | 9.85 | 6900 | 0 | 5 | 0 | 1 | 0 | 1 |
| **3** | 2.85 | 4.15 | 5200 | 0 | 11 | 0 | 1 | 0 | 1 |
| **4** | 4.60 | 6.87 | 42450 | 0 | 8 | 1 | 0 | 0 | 1 |

In [27]:

final\_dataset**.**corr()

# Out[27]: Selling\_Price Present\_Price Kms\_Driven Owner No.Year Fuel\_Type\_Diesel Fuel\_Type\_Petrol Seller\_Type\_Individual Transmission\_Manual

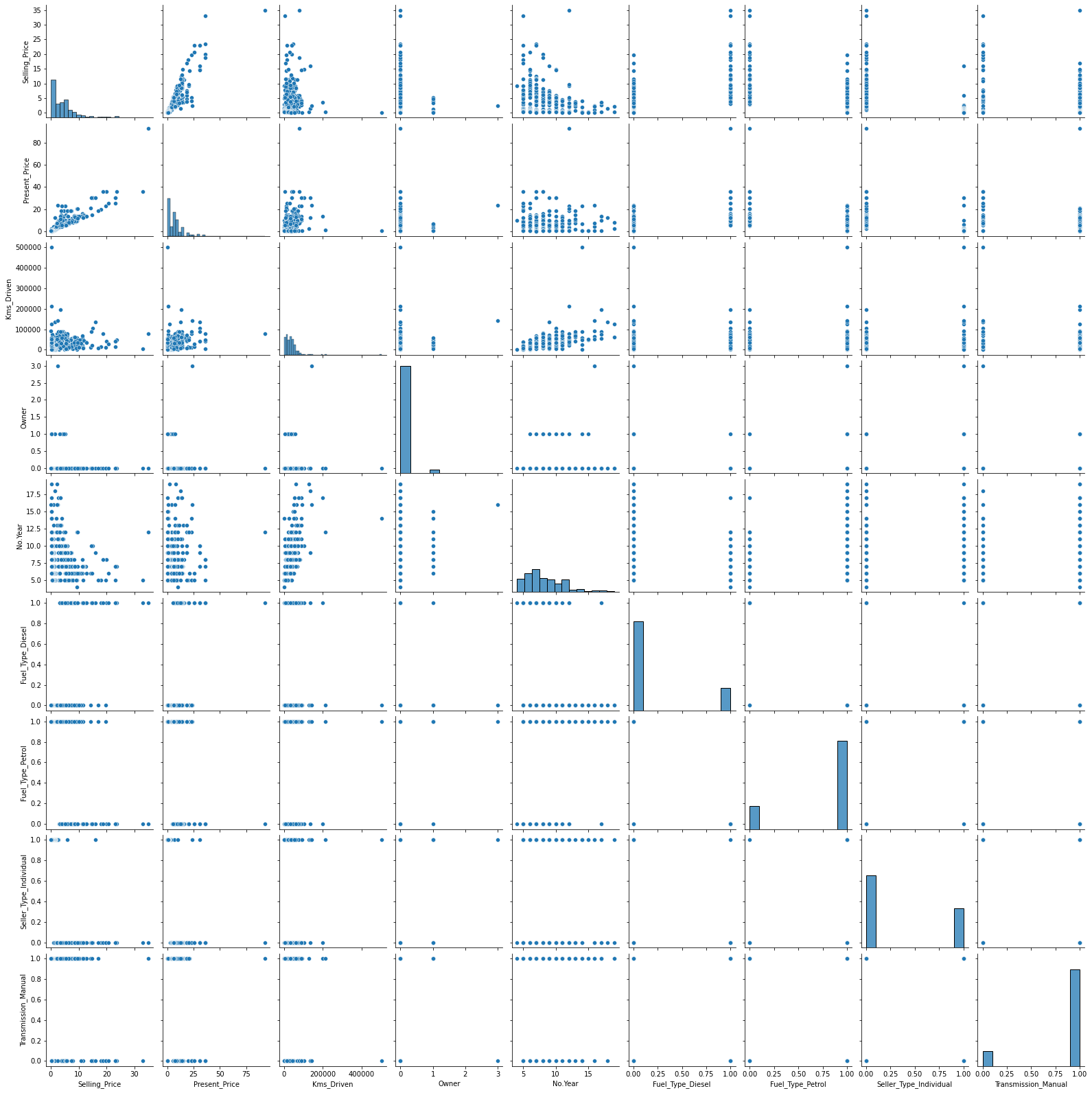
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Selling\_Price** | 1.000000 | 0.878983 | 0.029187 -0.088344 -0.236141 | 0.552339 | -0.540571 | -0.550724 | -0.367128 |
| **Present\_Price** | 0.878983 | 1.000000 | 0.203647 0.008057 0.047584 | 0.473306 | -0.465244 | -0.512030 | -0.348715 |
| **Kms\_Driven** | 0.029187 | 0.203647 | 1.000000 0.089216 0.524342 | 0.172515 | -0.172874 | -0.101419 | -0.162510 |
| **Owner** | -0.088344 | 0.008057 | 0.089216 1.000000 0.182104 | -0.053469 | 0.055687 | 0.124269 | -0.050316 |
| **No.Year** | -0.236141 | 0.047584 | 0.524342 0.182104 1.000000 | -0.064315 | 0.059959 | 0.039896 | -0.000394 |
| **Fuel\_Type\_Diesel** | 0.552339 | 0.473306 | 0.172515 -0.053469 -0.064315 | 1.000000 | -0.979648 | -0.350467 | -0.098643 |
| **Fuel\_Type\_Petrol** | -0.540571 | -0.465244 | -0.172874 0.055687 0.059959 | -0.979648 | 1.000000 | 0.358321 | 0.091013 |
| **Seller\_Type\_Individual** | -0.550724 | -0.512030 | -0.101419 0.124269 0.039896 | -0.350467 | 0.358321 | 1.000000 | 0.063240 |
| **Transmission\_Manual** | -0.367128 | -0.348715 | -0.162510 -0.050316 -0.000394 | -0.098643 | 0.091013 | 0.063240 | 1.000000 |

In [28]:

**import** seaborn **as** sns

In [29]:

sns**.**pairplot(final\_dataset)

Out[29]: <seaborn.axisgrid.PairGrid at 0x15e60c40ac0>

**import** matplotlib.pyplot **as** plt

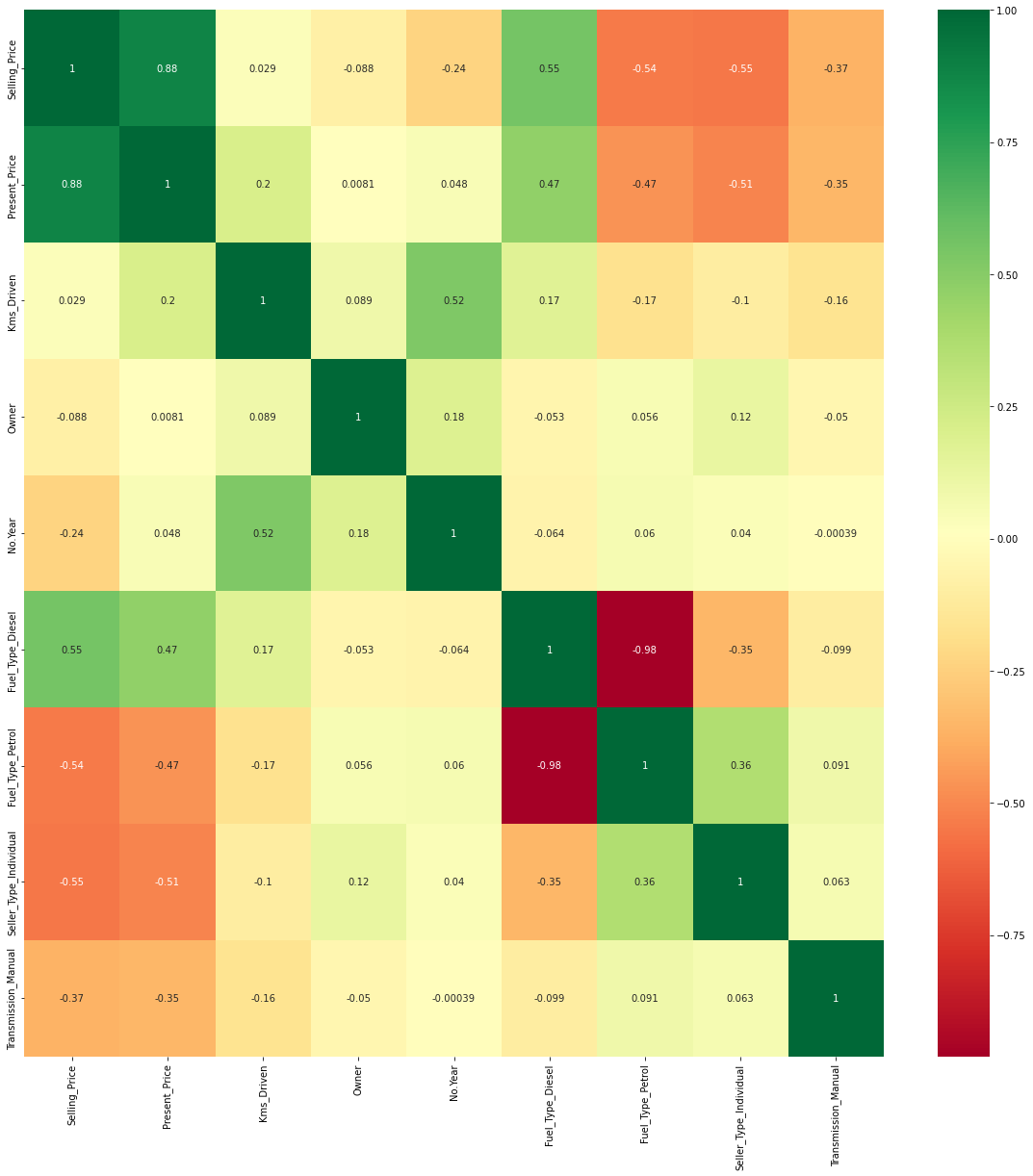
**%matplotlib** inline

In [30]:

In [31]:

corrmat**=**final\_dataset**.**corr() top\_corr\_features**=**corrmat**.**index plt**.**figure(figsize**=**(20,20))

g**=**sns**.**heatmap(final\_dataset[top\_corr\_features]**.**corr(),annot**=True**,cmap**=**"RdYlGn")



In [32]:

final\_dataset**.**head()

# Out[32]: Selling\_Price Present\_Price Kms\_Driven Owner No.Year Fuel\_Type\_Diesel Fuel\_Type\_Petrol Seller\_Type\_Individual Transmission\_Manual

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **0** | 3.35 | 5.59 | 27000 | 0 | 8 | 0 | 1 | 0 | 1 |
| **1** | 4.75 | 9.54 | 43000 | 0 | 9 | 1 | 0 | 0 | 1 |
| **2** | 7.25 | 9.85 | 6900 | 0 | 5 | 0 | 1 | 0 | 1 |
| **3** | 2.85 | 4.15 | 5200 | 0 | 11 | 0 | 1 | 0 | 1 |
| **4** | 4.60 | 6.87 | 42450 | 0 | 8 | 1 | 0 | 0 | 1 |

In [33]:

X**=**final\_dataset**.**iloc[:,1:] y**=**final\_dataset**.**iloc[:,0]

In [34]:

X**.**head()

Out[34]:

# Present\_Price Kms\_Driven Owner No.Year Fuel\_Type\_Diesel Fuel\_Type\_Petrol Seller\_Type\_Individual Transmission\_Manual 0 5.59 27000 0 8 0 1 0 1

**2** 9.85 6900 0 5 0 1 0 1

**3** 4.15 5200 0 11 0 1 0 1

**1** 9.54 43000 0 9 1 0 0 1

**4** 6.87 42450 0 8 1 0 0 1

In [35]:

y**.**head()

Out[35]: 0 3.35

1 4.75

2 7.25

3 2.85

4 4.60

Name: Selling\_Price, dtype: float64