

# car prediction

May 10, 2021

```
[2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stat
import warnings
warnings.filterwarnings('ignore')
```

```
C:\Users\Chandra Sekhar\Anaconda3\lib\site-
packages\statsmodels\tools\_testing.py:19: FutureWarning: pandas.util.testing is
deprecated. Use the functions in the public API at pandas.testing instead.
import pandas.util.testing as tm
```

```
[3]: df=pd.read_csv('car data.csv')
```

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

```
[4]: Car_Name  Year  Selling_Price  Present_Price  Kms_Driven  Fuel_Type  \
0    ritz    2014         3.35         5.59        27000    Petrol
1    sx4    2013         4.75         9.54        43000    Diesel
2    ciaz    2017         7.25         9.85         6900    Petrol
3  wagon r    2011         2.85         4.15         5200    Petrol
4    swift    2014         4.60         6.87        42450    Diesel
```

```
      Seller_Type  Transmission  Owner
0      Dealer      Manual      0
1      Dealer      Manual      0
2      Dealer      Manual      0
3      Dealer      Manual      0
4      Dealer      Manual      0
```

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

```
[5]: (301, 9)
```

```
[6]: print(df['Seller_Type'].unique())
```

```
['Dealer' 'Individual']
```

```
[7]: print(df['Transmission'].unique())
```

```
['Manual' 'Automatic']
```

```
[8]: print(df['Owner'].unique())
```

```
[0 1 3]
```

```
[9]: df.isnull().sum()
```

```
[9]: 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
```

```
[10]: df.describe()
```

```
[10]:
```

	Year	Selling_Price	Present_Price	Kms_Driven	Owner
count	301.000000	301.000000	301.000000	301.000000	301.000000
mean	2013.627907	4.661296	7.628472	36947.205980	0.043189
std	2.891554	5.082812	8.644115	38886.883882	0.247915
min	2003.000000	0.100000	0.320000	500.000000	0.000000
25%	2012.000000	0.900000	1.200000	15000.000000	0.000000
50%	2014.000000	3.600000	6.400000	32000.000000	0.000000
75%	2016.000000	6.000000	9.900000	48767.000000	0.000000
max	2018.000000	35.000000	92.600000	500000.000000	3.000000

```
[12]: df.columns
```

```
[12]: Index(['Car_Name', 'Year', 'Selling_Price', 'Present_Price', 'Kms_Driven',
        'Fuel_Type', 'Seller_Type', 'Transmission', 'Owner'],
        dtype='object')
```

```
[13]: df['current_year']=2021
```

```
[14]: df.head()
```

```
[14]:
```

	Car_Name	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	\
0	ritz	2014	3.35	5.59	27000	Petrol	
1	sx4	2013	4.75	9.54	43000	Diesel	
2	ciaz	2017	7.25	9.85	6900	Petrol	
3	wagon r	2011	2.85	4.15	5200	Petrol	
4	swift	2014	4.60	6.87	42450	Diesel	

```
Seller_Type Transmission Owner current_year
```

0	Dealer	Manual	0	2021
1	Dealer	Manual	0	2021
2	Dealer	Manual	0	2021
3	Dealer	Manual	0	2021
4	Dealer	Manual	0	2021

```
[15]: df.columns
```

```
[15]: Index(['Car_Name', 'Year', 'Selling_Price', 'Present_Price', 'Kms_Driven',
        'Fuel_Type', 'Seller_Type', 'Transmission', 'Owner', 'current_year'],
        dtype='object')
```

```
[16]: no_year=df['current_year']-df['Year']
```

```
[17]: df['n_years']=no_year
```

```
[18]: df.head()
```

```
[18]: Car_Name  Year  Selling_Price  Present_Price  Kms_Driven  Fuel_Type  \
0      ritz  2014           3.35           5.59       27000    Petrol
1      sx4   2013           4.75           9.54       43000    Diesel
2      ciaz  2017           7.25           9.85        6900    Petrol
3  wagon r  2011           2.85           4.15        5200    Petrol
4    swift  2014           4.60           6.87      42450    Diesel

      Seller_Type  Transmission  Owner  current_year  n_years
0      Dealer      Manual      0      2021         7
1      Dealer      Manual      0      2021         8
2      Dealer      Manual      0      2021         4
3      Dealer      Manual      0      2021        10
4      Dealer      Manual      0      2021         7
```

```
[19]: df.drop('current_year',axis=1,inplace=True)
```

```
[20]: df.head()
```

```
[20]: Car_Name  Year  Selling_Price  Present_Price  Kms_Driven  Fuel_Type  \
0      ritz  2014           3.35           5.59       27000    Petrol
1      sx4   2013           4.75           9.54       43000    Diesel
2      ciaz  2017           7.25           9.85        6900    Petrol
3  wagon r  2011           2.85           4.15        5200    Petrol
4    swift  2014           4.60           6.87      42450    Diesel

      Seller_Type  Transmission  Owner  n_years
0      Dealer      Manual      0         7
1      Dealer      Manual      0         8
2      Dealer      Manual      0         4
3      Dealer      Manual      0        10
4      Dealer      Manual      0         7
```

```
[30]: df.columns
```

```
[30]: Index(['Year', 'Selling_Price', 'Present_Price', 'Kms_Driven', 'Fuel_Type',
          'Seller_Type', 'Transmission', 'Owner', 'n_years'],
          dtype='object')
```

```
[ ]:
```

```
[31]: df1=pd.get_dummies(df,columns=['Fuel_Type','Seller_Type','Transmission'])
```

```
[32]: df1.head()
```

```
[32]:   Year  Selling_Price  Present_Price  Kms_Driven  Owner  n_years  \
0  2014             3.35           5.59       27000      0         7
1  2013             4.75           9.54       43000      0         8
2  2017             7.25           9.85        6900      0         4
3  2011             2.85           4.15        5200      0        10
4  2014             4.60           6.87       42450      0         7
```

```
      Fuel_Type_CNG  Fuel_Type_Diesel  Fuel_Type_Petrol  Seller_Type_Dealer  \
0                0                0                1                1
1                0                1                0                1
2                0                0                1                1
3                0                0                1                1
4                0                1                0                1
```

```
      Seller_Type_Individual  Transmission_Automatic  Transmission_Manual
0                0                0                1
1                0                0                1
2                0                0                1
3                0                0                1
4                0                0                1
```

```
[33]: df1.describe()
```

```
[33]:   count      Year  Selling_Price  Present_Price  Kms_Driven  Owner  \
count    301.000000    301.000000    301.000000    301.000000    301.000000
mean     2013.627907     4.661296     7.628472    36947.205980     0.043189
std        2.891554     5.082812     8.644115    38886.883882     0.247915
min       2003.000000     0.100000     0.320000     500.000000     0.000000
25%       2012.000000     0.900000     1.200000    15000.000000     0.000000
50%       2014.000000     3.600000     6.400000    32000.000000     0.000000
75%       2016.000000     6.000000     9.900000    48767.000000     0.000000
max       2018.000000    35.000000    92.600000   500000.000000     3.000000
```

```
      n_years  Fuel_Type_CNG  Fuel_Type_Diesel  Fuel_Type_Petrol  \
count    301.000000    301.000000    301.000000    301.000000
mean        7.372093     0.006645     0.199336     0.794020
std        2.891554     0.081378     0.400166     0.405089
min         3.000000     0.000000     0.000000     0.000000
25%         5.000000     0.000000     0.000000     1.000000
50%         7.000000     0.000000     0.000000     1.000000
```

75%	9.000000	0.000000	0.000000	1.000000
max	18.000000	1.000000	1.000000	1.000000

	Seller_Type_Dealer	Seller_Type_Individual	Transmission_Automatic	\
count	301.000000	301.000000	301.000000	
mean	0.647841	0.352159	0.132890	
std	0.478439	0.478439	0.340021	
min	0.000000	0.000000	0.000000	
25%	0.000000	0.000000	0.000000	
50%	1.000000	0.000000	0.000000	
75%	1.000000	1.000000	0.000000	
max	1.000000	1.000000	1.000000	

	Transmission_Manual
count	301.000000
mean	0.867110
std	0.340021
min	0.000000
25%	1.000000
50%	1.000000
75%	1.000000
max	1.000000

```
[34]: df1.corr()
```

```
[34]:
```

	Year	Selling_Price	Present_Price	Kms_Driven	\
Year	1.000000	0.236141	-0.047584	-0.524342	
Selling_Price	0.236141	1.000000	0.878983	0.029187	
Present_Price	-0.047584	0.878983	1.000000	0.203647	
Kms_Driven	-0.524342	0.029187	0.203647	1.000000	
Owner	-0.182104	-0.088344	0.008057	0.089216	
n_years	-1.000000	-0.236141	0.047584	0.524342	
Fuel_Type_CNG	-0.017790	-0.025164	-0.011500	0.012223	
Fuel_Type_Diesel	0.064315	0.552339	0.473306	0.172515	
Fuel_Type_Petrol	-0.059959	-0.540571	-0.465244	-0.172874	
Seller_Type_Dealer	0.039896	0.550724	0.512030	0.101419	
Seller_Type_Individual	-0.039896	-0.550724	-0.512030	-0.101419	
Transmission_Automatic	-0.000394	0.367128	0.348715	0.162510	
Transmission_Manual	0.000394	-0.367128	-0.348715	-0.162510	

	Owner	n_years	Fuel_Type_CNG	Fuel_Type_Diesel	\
Year	-0.182104	-1.000000	-0.017790	0.064315	
Selling_Price	-0.088344	-0.236141	-0.025164	0.552339	
Present_Price	0.008057	0.047584	-0.011500	0.473306	
Kms_Driven	0.089216	0.524342	0.012223	0.172515	
Owner	1.000000	0.182104	-0.014272	-0.053469	
n_years	0.182104	1.000000	0.017790	-0.064315	
Fuel_Type_CNG	-0.014272	0.017790	1.000000	-0.040808	

Fuel_Type_Diesel	-0.053469	-0.064315	-0.040808	1.000000
Fuel_Type_Petrol	0.055687	0.059959	-0.160577	-0.979648
Seller_Type_Dealer	-0.124269	-0.039896	0.060300	0.350467
Seller_Type_Individual	0.124269	0.039896	-0.060300	-0.350467
Transmission_Automatic	0.050316	0.000394	-0.032018	0.098643
Transmission_Manual	-0.050316	-0.000394	0.032018	-0.098643

	Fuel_Type_Petrol	Seller_Type_Dealer \
Year	-0.059959	0.039896
Selling_Price	-0.540571	0.550724
Present_Price	-0.465244	0.512030
Kms_Driven	-0.172874	0.101419
Owner	0.055687	-0.124269
n_years	0.059959	-0.039896
Fuel_Type_CNG	-0.160577	0.060300
Fuel_Type_Diesel	-0.979648	0.350467
Fuel_Type_Petrol	1.000000	-0.358321
Seller_Type_Dealer	-0.358321	1.000000
Seller_Type_Individual	0.358321	-1.000000
Transmission_Automatic	-0.091013	0.063240
Transmission_Manual	0.091013	-0.063240

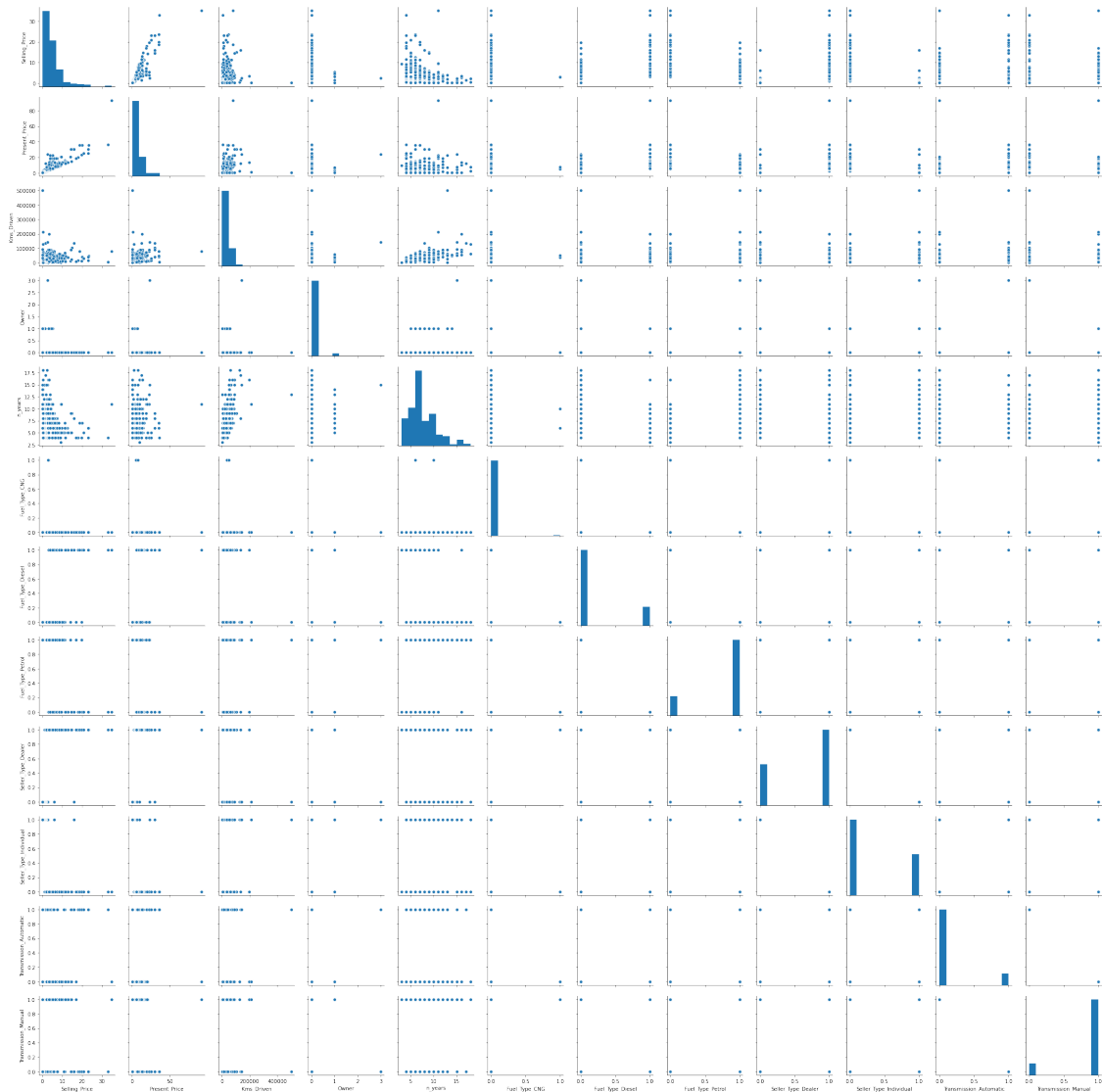
	Seller_Type_Individual	Transmission_Automatic \
Year	-0.039896	-0.000394
Selling_Price	-0.550724	0.367128
Present_Price	-0.512030	0.348715
Kms_Driven	-0.101419	0.162510
Owner	0.124269	0.050316
n_years	0.039896	0.000394
Fuel_Type_CNG	-0.060300	-0.032018
Fuel_Type_Diesel	-0.350467	0.098643
Fuel_Type_Petrol	0.358321	-0.091013
Seller_Type_Dealer	-1.000000	0.063240
Seller_Type_Individual	1.000000	-0.063240
Transmission_Automatic	-0.063240	1.000000
Transmission_Manual	0.063240	-1.000000

	Transmission_Manual
Year	0.000394
Selling_Price	-0.367128
Present_Price	-0.348715
Kms_Driven	-0.162510
Owner	-0.050316
n_years	-0.000394
Fuel_Type_CNG	0.032018
Fuel_Type_Diesel	-0.098643
Fuel_Type_Petrol	0.091013

Seller_Type_Dealer	-0.063240
Seller_Type_Individual	0.063240
Transmission_Automatic	-1.000000
Transmission_Manual	1.000000

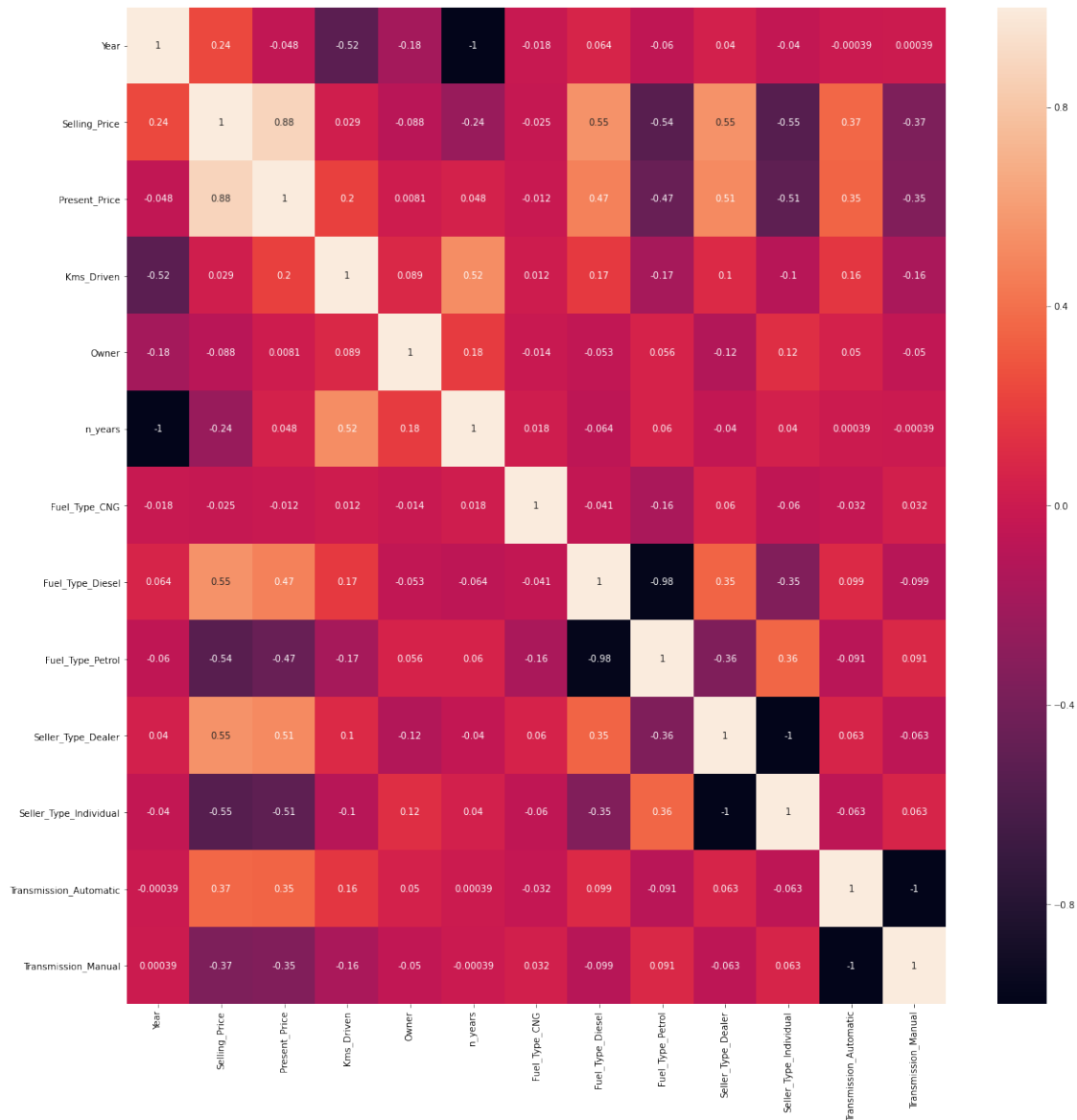
```
[52]: sns.pairplot(df1)
```

```
[52]: <seaborn.axisgrid.PairGrid at 0x1d5d620d9b0>
```



```
[35]: corr=df1.corr()
plt.figure(figsize=(20,20))
sns.heatmap(corr,annot=True)
```

```
[35]: <matplotlib.axes._subplots.AxesSubplot at 0x1d8f1c5b400>
```



```
[36]: df1.columns
```

```
[36]: Index(['Year', 'Selling_Price', 'Present_Price', 'Kms_Driven', 'Owner',
'n_years', 'Fuel_Type_CNG', 'Fuel_Type_Diesel', 'Fuel_Type_Petrol',
'Seller_Type_Dealer', 'Seller_Type_Individual',
'Transmission_Automatic', 'Transmission_Manual'],
dtype='object')
```

```
[ ]:
```

```
[37]: df1.head()
```

```
[37]:   Year  Selling_Price  Present_Price  Kms_Driven  Owner  n_years  \
0  2014             3.35           5.59       27000      0         7
```



1	2013	4.75	9.54	43000	0	8
2	2017	7.25	9.85	6900	0	4
3	2011	2.85	4.15	5200	0	10
4	2014	4.60	6.87	42450	0	7

	Fuel_Type_CNG	Fuel_Type_Diesel	Fuel_Type_Petrol	Seller_Type_Dealer	\
0	0	0	1		1
1	0	1	0		1
2	0	0	1		1
3	0	0	1		1
4	0	1	0		1

	Seller_Type_Individual	Transmission_Automatic	Transmission_Manual
0	0	0	1
1	0	0	1
2	0	0	1
3	0	0	1
4	0	0	1

```
[61]: x=df1.drop('Selling_Price',axis=1)
```

```
[62]: y=df1['Selling_Price']
```

```
[ ]:
```

```
[ ]: #when we deal with many number of features at correlation by using the some_  

→conditional formate and apply the threshold value
```

```
threshold=0.8
```

```
def correlation(df,threshold):
    set_col=set() #all columns in dataset
    corr=df.corr() #applying the correlation
    for i in range(len(corr.columns)):
        for j in range(i):
            if abs(corr.iloc[i,j]) > threshold:
                colname=corr.columns[i]
                set_col.add(colname)
            return set_col

    print(correlation)
```

```
[ ]: #outliers detection
```

```
def outliers(df,features)
outliers_indicate=[]
for i in features:
    Q1=np.percentaile(df[i].25)
    Q3=np.percentaile(df[i].75)

    IQR=Q3-Q1
```

```
outlier_setup=IQR*1.5
```

```
outlier_list_column=df(df[i]<Q1-outlier_step | df(df[i]> + outlier_step).  
→index)
```

```
outlierindicates.extended(outlie_list_column)
```

```
outlier_indicates=counter(outlier_indicates)
```

```
multiple_outlier_list=list(i for i v in outlier_indicate.items() if v>2)  
return mulitple_outliers
```

```
[72]: n_estimators=[int(x) for x in np.linspace(start=100,stop=2000,num=20)]  
max_features=['auto','sqrt','log2']  
max_depth=[int(x) for x in np.linspace(10,1000,10)]  
min_samples_split=[2,5,10,15,100]  
min_samples_leaf=[1,2,5,10]  
random_grid={  
    'n_estimators':n_estimators,  
    'max_features':max_features,  
    'max_depth':max_depth,  
    'min_samples_split':min_samples_split,  
    'min_samples_leaf':min_samples_leaf,  
    'criterion':['MSE']  
}
```

```
[ ]: random_grid
```

```
[64]: from sklearn.model_selection import train_test_split  
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.3,random_state=0)
```

```
[65]: from sklearn.ensemble import RandomForestRegressor
```

```
[66]: rf=RandomForestRegressor()
```

```
[67]: rf.fit(x_train,y_train)
```

```
[67]: RandomForestRegressor()
```

```
[68]: from sklearn.model_selection import RandomizedSearchCV
```

```
[79]: from sklearn.metrics import mean_squared_error
```

```
[92]: rcv=RandomizedSearchCV(estimator=rf,param_distributions=random_grid,scoring='neg_mean_squared_
```

```
[93]: rcv
```

```
[93]: RandomizedSearchCV(cv=3, estimator=RandomForestRegressor(), n_iter=100,  
                        n_jobs=1,  
                        param_distributions={'criterion': ['MSE'],  
                                             'max_depth': [10, 120, 230, 340, 450,  
                                                            560, 670, 780, 890,  
                                                            1000],  
                                             'max_features': ['auto', 'sqrt',  
                                                             'log2']},
```

```

        'min_samples_leaf': [1, 2, 5, 10],
        'min_samples_split': [2, 5, 10, 15,
                                100],
        'n_estimators': [100, 200, 300, 400,
                           500, 600, 700, 800,
                           900, 1000, 1100, 1200,
                           1300, 1400, 1500, 1600,
                           1700, 1800, 1900,
                           2000]}},
        random_state=100, scoring='neg_mean_squared_error',
        verbose=2)

```

[94]: `rcv.fit(x_train,y_train)`

```

Fitting 3 folds for each of 100 candidates, totalling 300 fits
[CV] n_estimators=100, min_samples_split=10, min_samples_leaf=1,
max_features=log2, max_depth=450, criterion=MSE
[CV] n_estimators=100, min_samples_split=10, min_samples_leaf=1,
max_features=log2, max_depth=450, criterion=MSE, total= 0.1s
[CV] n_estimators=100, min_samples_split=10, min_samples_leaf=1,
max_features=log2, max_depth=450, criterion=MSE
[CV] n_estimators=100, min_samples_split=10, min_samples_leaf=1,
max_features=log2, max_depth=450, criterion=MSE, total= 0.0s
[CV] n_estimators=100, min_samples_split=10, min_samples_leaf=1,
max_features=log2, max_depth=450, criterion=MSE
[CV] n_estimators=100, min_samples_split=10, min_samples_leaf=1,
max_features=log2, max_depth=450, criterion=MSE, total= 0.0s
[CV] n_estimators=1700, min_samples_split=5, min_samples_leaf=2,
max_features=log2, max_depth=560, criterion=MSE

[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
[Parallel(n_jobs=1)]: Done 1 out of 1 | elapsed: 0.0s remaining: 0.0s

[CV] n_estimators=1700, min_samples_split=5, min_samples_leaf=2,
max_features=log2, max_depth=560, criterion=MSE, total= 0.7s
[CV] n_estimators=1700, min_samples_split=5, min_samples_leaf=2,
max_features=log2, max_depth=560, criterion=MSE
[CV] n_estimators=1700, min_samples_split=5, min_samples_leaf=2,
max_features=log2, max_depth=560, criterion=MSE, total= 0.7s
[CV] n_estimators=1700, min_samples_split=5, min_samples_leaf=2,
max_features=log2, max_depth=560, criterion=MSE
[CV] n_estimators=1700, min_samples_split=5, min_samples_leaf=2,
max_features=log2, max_depth=560, criterion=MSE, total= 0.8s
[CV] n_estimators=2000, min_samples_split=5, min_samples_leaf=1,
max_features=log2, max_depth=670, criterion=MSE
[CV] n_estimators=2000, min_samples_split=5, min_samples_leaf=1,
max_features=log2, max_depth=670, criterion=MSE, total= 0.9s
[CV] n_estimators=2000, min_samples_split=5, min_samples_leaf=1,

```





















[illegible]



[illegible]

```
max_features=sqrt, max_depth=1000, criterion=MSE  
[CV] n_estimators=800, min_samples_split=15, min_samples_leaf=10,  
max_features=sqrt, max_depth=1000, criterion=MSE, total=    0.3s  
[CV] n_estimators=800, min_samples_split=15, min_samples_leaf=10,  
max_features=sqrt, max_depth=1000, criterion=MSE  
[CV] n_estimators=800, min_samples_split=15, min_samples_leaf=10,  
max_features=sqrt, max_depth=1000, criterion=MSE, total=    0.4s  
[CV] n_estimators=1000, min_samples_split=10, min_samples_leaf=10,  
max_features=auto, max_depth=340, criterion=MSE  
[CV] n_estimators=1000, min_samples_split=10, min_samples_leaf=10,  
max_features=auto, max_depth=340, criterion=MSE, total=    0.4s  
[CV] n_estimators=1000, min_samples_split=10, min_samples_leaf=10,  
max_features=auto, max_depth=340, criterion=MSE  
[CV] n_estimators=1000, min_samples_split=10, min_samples_leaf=10,  
max_features=auto, max_depth=340, criterion=MSE, total=    0.4s  
[CV] n_estimators=1000, min_samples_split=10, min_samples_leaf=10,  
max_features=auto, max_depth=340, criterion=MSE  
[CV] n_estimators=700, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=10, criterion=MSE  
[CV] n_estimators=700, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=10, criterion=MSE, total=    0.3s  
[CV] n_estimators=700, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=10, criterion=MSE  
[CV] n_estimators=700, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=10, criterion=MSE, total=    0.3s  
[CV] n_estimators=700, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=10, criterion=MSE  
[CV] n_estimators=700, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=10, criterion=MSE, total=    0.3s  
[CV] n_estimators=1300, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=890, criterion=MSE  
[CV] n_estimators=1300, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=890, criterion=MSE, total=    0.5s  
[CV] n_estimators=1300, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=890, criterion=MSE  
[CV] n_estimators=1300, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=890, criterion=MSE, total=    0.5s  
[CV] n_estimators=1300, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=890, criterion=MSE  
[CV] n_estimators=1300, min_samples_split=2, min_samples_leaf=1,  
max_features=log2, max_depth=890, criterion=MSE, total=    0.5s  
[CV] n_estimators=700, min_samples_split=15, min_samples_leaf=1,  
max_features=auto, max_depth=780, criterion=MSE  
[CV] n_estimators=700, min_samples_split=15, min_samples_leaf=1,  
max_features=auto, max_depth=780, criterion=MSE, total=    0.3s  
[CV] n_estimators=700, min samples split=15, min samples leaf=1,
```





[illegible]



[illegible]



```

max_features=sqrt, max_depth=120, criterion=MSE
[CV] n_estimators=800, min_samples_split=15, min_samples_leaf=1,
max_features=sqrt, max_depth=120, criterion=MSE, total= 0.4s
[CV] n_estimators=800, min_samples_split=15, min_samples_leaf=1,
max_features=sqrt, max_depth=120, criterion=MSE
[CV] n_estimators=800, min_samples_split=15, min_samples_leaf=1,
max_features=sqrt, max_depth=120, criterion=MSE, total= 0.4s
[CV] n_estimators=900, min_samples_split=5, min_samples_leaf=5,
max_features=log2, max_depth=780, criterion=MSE
[CV] n_estimators=900, min_samples_split=5, min_samples_leaf=5,
max_features=log2, max_depth=780, criterion=MSE, total= 0.4s
[CV] n_estimators=900, min_samples_split=5, min_samples_leaf=5,
max_features=log2, max_depth=780, criterion=MSE
[CV] n_estimators=900, min_samples_split=5, min_samples_leaf=5,
max_features=log2, max_depth=780, criterion=MSE, total= 0.4s
[CV] n_estimators=900, min_samples_split=5, min_samples_leaf=5,
max_features=log2, max_depth=780, criterion=MSE
[CV] n_estimators=700, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=780, criterion=MSE
[CV] n_estimators=700, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=780, criterion=MSE, total= 0.4s
[CV] n_estimators=700, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=780, criterion=MSE
[CV] n_estimators=700, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=780, criterion=MSE, total= 0.3s
[CV] n_estimators=700, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=780, criterion=MSE
[CV] n_estimators=700, min_samples_split=100, min_samples_leaf=2,
max_features=auto, max_depth=780, criterion=MSE, total= 0.3s
[CV] n_estimators=700, min_samples_split=15, min_samples_leaf=2,
max_features=log2, max_depth=1000, criterion=MSE
[CV] n_estimators=700, min_samples_split=15, min_samples_leaf=2,
max_features=log2, max_depth=1000, criterion=MSE, total= 0.3s
[CV] n_estimators=700, min_samples_split=15, min_samples_leaf=2,
max_features=log2, max_depth=1000, criterion=MSE
[CV] n_estimators=700, min_samples_split=15, min_samples_leaf=2,
max_features=log2, max_depth=1000, criterion=MSE, total= 0.3s
[CV] n_estimators=700, min_samples_split=15, min_samples_leaf=2,
max_features=log2, max_depth=1000, criterion=MSE
[CV] n_estimators=1200, min_samples_split=100, min_samples_leaf=2,
max_features=log2, max_depth=340, criterion=MSE
[CV] n_estimators=1200, min_samples_split=100, min_samples_leaf=2,
max_features=log2, max_depth=340, criterion=MSE, total= 0.5s
[CV] n_estimators=1200, min_samples_split=100, min_samples_leaf=2,

```









[illegible]



```

max_features=sqrt, max_depth=890, criterion=MSE
[CV] n_estimators=1800, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=890, criterion=MSE, total= 0.9s
[CV] n_estimators=1800, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=890, criterion=MSE
[CV] n_estimators=1800, min_samples_split=5, min_samples_leaf=1,
max_features=sqrt, max_depth=890, criterion=MSE, total= 0.8s
[CV] n_estimators=1500, min_samples_split=2, min_samples_leaf=2,
max_features=auto, max_depth=450, criterion=MSE
[CV] n_estimators=1500, min_samples_split=2, min_samples_leaf=2,
max_features=auto, max_depth=450, criterion=MSE, total= 0.8s
[CV] n_estimators=1500, min_samples_split=2, min_samples_leaf=2,
max_features=auto, max_depth=450, criterion=MSE
[CV] n_estimators=1500, min_samples_split=2, min_samples_leaf=2,
max_features=auto, max_depth=450, criterion=MSE, total= 0.7s
[CV] n_estimators=1500, min_samples_split=2, min_samples_leaf=2,
max_features=auto, max_depth=450, criterion=MSE
[CV] n_estimators=1500, min_samples_split=2, min_samples_leaf=2,
max_features=auto, max_depth=450, criterion=MSE, total= 0.7s

[Parallel(n_jobs=1)]: Done 300 out of 300 | elapsed: 2.3min finished

```

```

└─
↳-----

KeyError                                Traceback (most recent call└
↳last)

<ipython-input-94-b757a035f514> in <module>
----> 1 rcv.fit(x_train,y_train)

~\Anaconda3\lib\site-packages\sklearn\utils\validation.py in└
↳inner_f(*args, **kwargs)
    70                                     FutureWarning)
    71         kwargs.update({k: arg for k, arg in zip(sig.parameters,└
↳args)})
--> 72         return f(**kwargs)
    73     return inner_f
    74

~\Anaconda3\lib\site-packages\sklearn\model_selection\_search.py in└
↳fit(self, X, y, groups, **fit_params)
    763         refit_start_time = time.time()
    764         if y is not None:
--> 765             self.best_estimator_.fit(X, y, **fit_params)

```

```

766             else:
767                 self.best_estimator_.fit(X, **fit_params)

~\Anaconda3\lib\site-packages\sklearn\ensemble\_forest.py in fit(self,
↳X, y, sample_weight)
    390                 verbose=self.verbose, class_weight=self.
↳class_weight,
    391                 n_samples_bootstrap=n_samples_bootstrap)
--> 392             for i, t in enumerate(trees))
    393
    394             # Collect newly grown trees

~\Anaconda3\lib\site-packages\joblib\parallel.py in __call__(self,
↳iterable)
    919             # remaining jobs.
    920             self._iterating = False
--> 921             if self.dispatch_one_batch(iterator):
    922                 self._iterating = self._original_iterator is not None
    923

~\Anaconda3\lib\site-packages\joblib\parallel.py in
↳dispatch_one_batch(self, iterator)
    757             return False
    758             else:
--> 759                 self._dispatch(tasks)
    760                 return True
    761

~\Anaconda3\lib\site-packages\joblib\parallel.py in _dispatch(self,
↳batch)
    714             with self._lock:
    715                 job_idx = len(self._jobs)
--> 716                 job = self._backend.apply_async(batch, callback=cb)
    717                 # A job can complete so quickly than its callback is
    718                 # called before we get here, causing self._jobs to

~\Anaconda3\lib\site-packages\joblib\_parallel_backends.py in
↳apply_async(self, func, callback)
    180         def apply_async(self, func, callback=None):
    181             """Schedule a func to be run"""
--> 182             result = ImmediateResult(func)
    183             if callback:

```

```
184         callback(result)
```

```
~\Anaconda3\lib\site-packages\joblib\_parallel_backends.py in
↳ __init__(self, batch)
    547         # Don't delay the application, to avoid keeping the input
    548         # arguments in memory
--> 549         self.results = batch()
    550
    551     def get(self):
```

```
~\Anaconda3\lib\site-packages\joblib\parallel.py in __call__(self)
    223         with parallel_backend(self._backend, n_jobs=self._n_jobs):
    224             return [func(*args, **kwargs)
--> 225                     for func, args, kwargs in self.items]
    226
    227     def __len__(self):
```

```
~\Anaconda3\lib\site-packages\joblib\parallel.py in <listcomp>(.0)
    223         with parallel_backend(self._backend, n_jobs=self._n_jobs):
    224             return [func(*args, **kwargs)
--> 225                     for func, args, kwargs in self.items]
    226
    227     def __len__(self):
```

```
~\Anaconda3\lib\site-packages\sklearn\ensemble\_forest.py in
↳ _parallel_build_trees(tree, forest, X, y, sample_weight, tree_idx, n_trees,
↳ verbose, class_weight, n_samples_bootstrap)
    166
↳ indices=indices)
    167
--> 168         tree.fit(X, y, sample_weight=curr_sample_weight,
↳ check_input=False)
    169     else:
    170         tree.fit(X, y, sample_weight=sample_weight,
↳ check_input=False)
```

```
~\Anaconda3\lib\site-packages\sklearn\tree\_classes.py in fit(self, X,
↳ y, sample_weight, check_input, X_idx_sorted)
    1244         sample_weight=sample_weight,
    1245         check_input=check_input,
-> 1246         X_idx_sorted=X_idx_sorted)
    1247     return self
```

1248

```
~\Anaconda3\lib\site-packages\sklearn\tree\_classes.py in fit(self, X,
->y, sample_weight, check_input, X_idx_sorted)
    334                                     self.
->n_classes_)
    335             else:
--> 336                 criterion = CRITERIA_REG[self.criterion](self.
->n_outputs_,
    337                                     n_samples)
    338
```

KeyError: 'MSE'

```
[ ]: predict=rcv.predict(x_test)
```

```
[ ]: sns.distplot(y_test-predict)
```

```
[ ]: plt.scatter(y_test,predict)
```

```
[100]: import pickle
        #open the file were you want to store the data
        file=open('car_price_prediction.pkl','wb')
        #dump informatin to that file
        pickle.dump(rcv,file)
```

```
[ ]:
```