CAR PRICE PREDICTION MODEL

```
#Loading dataset
import pandas as pd


data=pd.read_csv("/car data.csv")


#Top 5 rows of dataset
data.head()
```

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

```
# Number of rows and columns in dataset
data.shape
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
# Unique values in column sellertype
print(data['Seller_Type'].unique())
```

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

```
# Unique values in column Transmission
print(data['Transmission'].unique())
```

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

```
# Unique values in column Owner
print(data['Owner'].unique())
```

```
[0 1 3]
```

```
#Checking null values in dataset
data.isnull().sum()
```

✓ 0s completed at 12:02 PM

```
Year             0
Selling_Price    0
Present_Price    0
Kms_Driven       0
Fuel_Type        0
Seller_Type      0
Transmission     0
Owner            0
dtype: int64
```

```
#Overall stats of the dataset
data.describe()
```

	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

Automatic saving failed. This file was updated remotely or in another tab. Show diff

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

```
#Keeping only useful columns
final_dataset=data[['Year', 'Selling_Price', 'Present_Price', 'Kms_Driven',
       'Fuel_Type', 'Seller_Type', 'Transmission', 'Owner']]
```

```
#Top 5 rows of new dataset
final_dataset.head()
```

	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmissi
0	2014	3.35	5.59	27000	Petrol	Dealer	Manu
1	2013	4.75	9.54	43000	Diesel	Dealer	Manu
2	2017	7.25	9.85	6900	Petrol	Dealer	Manu

3	2011	2.85	4.15	5200	Petrol	Dealer	Manu
4	2014	4.60	6.87	42450	Diesel	Dealer	Manu

```
#Adding new column of current year i.e. 2023
final_dataset['Current_year']=2023
final_dataset.head()
```

	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmissi
0	2014	3.35	5.59	27000	Petrol	Dealer	Manu
1	2013	4.75	9.54	43000	Diesel	Dealer	Manu
2	2017	7.25	9.85	6900	Petrol	Dealer	Manu
3	2011	2.85	4.15	5200	Petrol	Dealer	Manu
4	2014	4.60	6.87	42450	Diesel	Dealer	Manu

```
#Adding column of age of the vehicle
final_dataset['Age']= final_dataset['Current_year']-final_dataset['Year']
final_dataset.head()
```

	Year	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmissi
0	2014	3.35	5.59	27000	Petrol	Dealer	Manu
						ler	Manu
						ler	Manu
3	2011	2.85	4.15	5200	Petrol	Dealer	Manu
4	2014	4.60	6.87	42450	Diesel	Dealer	Manu

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
#Removing columns of Year and Current year
final_dataset.drop(['Year'],axis=1,inplace=True)
final_dataset.drop(['Current_year'],axis=1,inplace=True)
final_dataset
```

	Selling_Price	Present_Price	Kms_Driven	Fuel_Type	Seller_Type	Transmission
0	3.35	5.59	27000	Petrol	Dealer	Manual
1	4.75	9.54	43000	Diesel	Dealer	Manual
2	7.25	9.85	6900	Petrol	Dealer	Manual
3	2.85	4.15	5200	Petrol	Dealer	Manual
4	4.60	6.87	42450	Diesel	Dealer	Manual

4	4.60	6.87	42450	Diesel	Dealer	Manual
...	...	...	...	...	...	...
296	9.50	11.60	33988	Diesel	Dealer	Manual
297	4.00	5.90	60000	Petrol	Dealer	Manual
298	3.35	11.00	87934	Petrol	Dealer	Manual
299	11.50	12.50	9000	Diesel	Dealer	Manual
300	5.30	5.90	5464	Petrol	Dealer	Manual

301 rows × 8 columns

```
final_dataset=pd.get_dummies(final_dataset,drop_first=True)
final_dataset.head()
```

	Selling_Price	Present_Price	Kms_Driven	Owner	Age	Fuel_Type_Diesel	Fuel_Type_
0	3.35	5.59	27000	0	9	0	
1	4.75	9.54	43000	0	10	1	
2	7.25	9.85	6900	0	6	0	
3	2.85	4.15	5200	0	12	0	
4	4.60	6.87	42450	0	9	1	

Automatic saving failed. This file was updated remotely or in another tab. Show diff

	Selling_Price	Present_Price	Kms_Driven	Owner	Age	F
Selling_Price	1.000000	0.878983	0.029187	-0.088344	-0.236141	
Present_Price	0.878983	1.000000	0.203647	0.008057	0.047584	
Kms_Driven	0.029187	0.203647	1.000000	0.089216	0.524342	
Owner	-0.088344	0.008057	0.089216	1.000000	0.182104	
Age	-0.236141	0.047584	0.524342	0.182104	1.000000	
Fuel_Type_Diesel	0.552339	0.473306	0.172515	-0.053469	-0.064315	
Fuel_Type_Petrol	-0.540571	-0.465244	-0.172874	0.055687	0.059959	
Seller_Type_Individual	-0.550724	-0.512030	-0.101419	0.124269	0.039896	
Transmission_Manual	-0.367128	-0.348715	-0.162510	-0.050316	-0.000394	

```
#Plotting the graph
import matplotlib.pyplot as plt
import seaborn as sns
sns.pairplot(final_dataset)
plt.show()
```

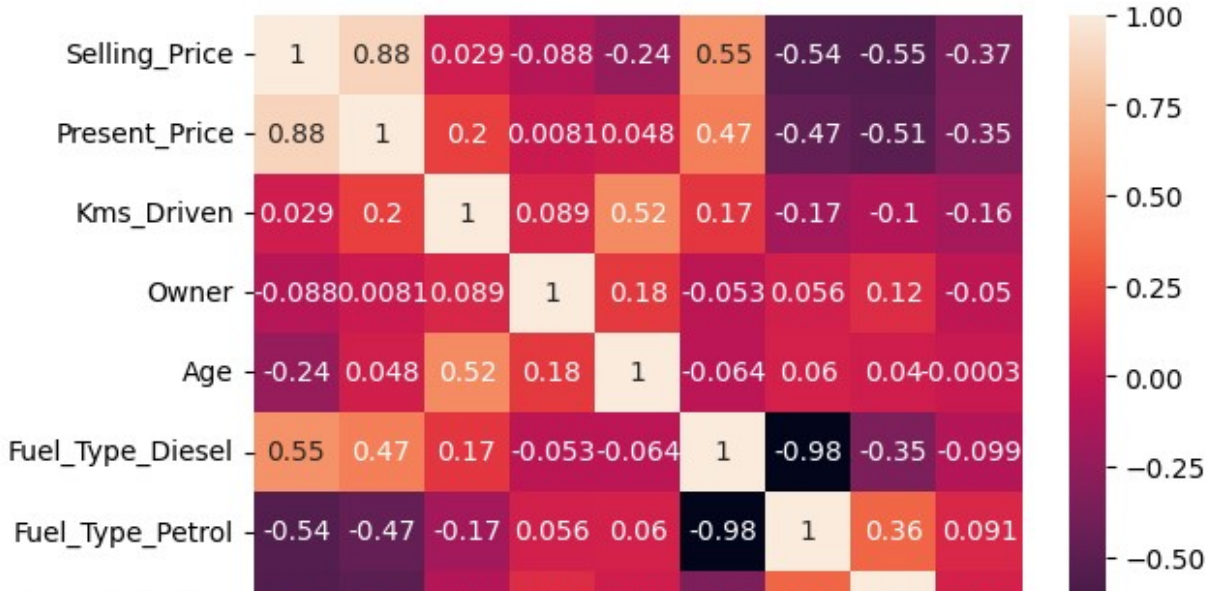
Automatic saving failed. This file was updated remotely or in another tab. Show diff

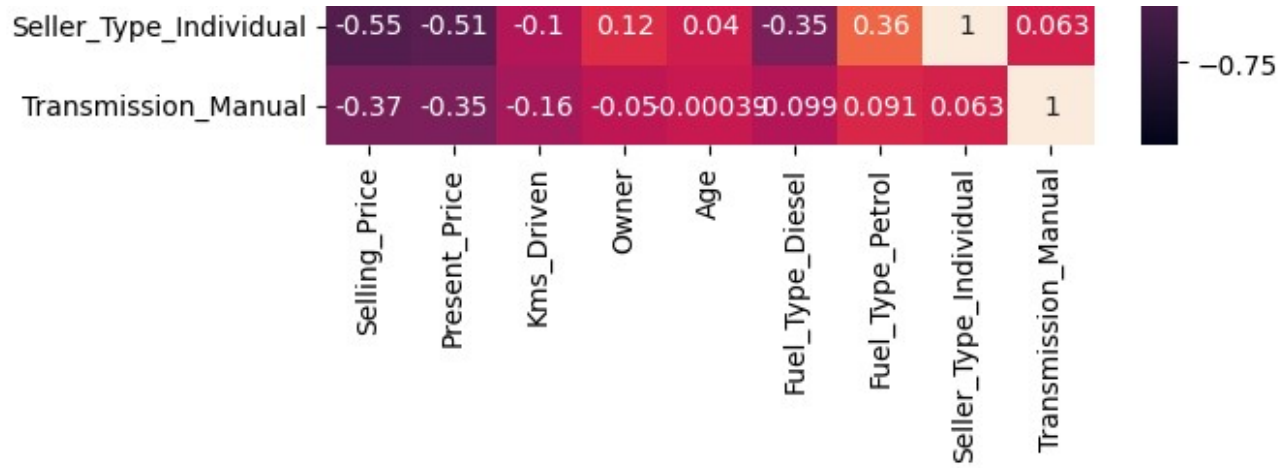
```
corrmat=final_dataset.corr()
top_corr_features=corrmat.index
plt.figure(figsize=(20,20))
plt.show()
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
sns.heatmap(final_dataset[top_corr_features].corr(),annot=True)
```

```
<Axes: >
```






```
final_dataset.head()
```

	Selling_Price	Present_Price	Kms_Driven	Owner	Age	Fuel_Type_Diesel	Fuel_Type_
0	3.35	5.59	27000	0	9	0	
1	4.75	9.54	43000	0	10	1	
2	7.25	9.85	6900	0	6	0	
3	2.85	4.15	5200	0	12	0	
4	4.60	6.87	42450	0	9	1	

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
y=final_dataset.iloc[:,0]
```

```
x.head()
```

	Present_Price	Kms_Driven	Owner	Age	Fuel_Type_Diesel	Fuel_Type_Petrol	Seller_
0	5.59	27000	0	9	0	1	
1	9.54	43000	0	10	1	0	
2	9.85	6900	0	6	0	1	
3	4.15	5200	0	12	0	1	
4	6.87	42450	0	9	1	0	

```
y.head()
```

```
0    3.35
1    4.75
```

```
2    7.25
3    2.85
4    4.60
Name: Selling_Price, dtype: float64
```

```
#Feature importance
from sklearn.ensemble import ExtraTreesRegressor
model=ExtraTreesRegressor()
model.fit(x,y)
```

```
▾ ExtraTreesRegressor
ExtraTreesRegressor()
```

```
print(model.feature_importances_)
```

```
[3.94534129e-01 3.97325881e-02 3.78191495e-04 7.64792950e-02
 2.17063693e-01 1.62550636e-02 1.30193389e-01 1.25363650e-01]
```

```
#Splitting
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2)
```

```
x_test.head(1)
```

	Present_Price	Kms_Driven	Owner	Age	Fuel_Type_Diesel	Fuel_Type_Petrol	Seller
							0

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
x_train.shape
```

```
(240, 8)
```

```
#Random Forest Regressor
from sklearn.ensemble import RandomForestRegressor
rf_random=RandomForestRegressor()
```

```
#Hyperperameters
import numpy as np
n_estimators=[int(x) for x in np.linspace(start=100,stop=1200,num=12)]
print(n_estimators)
```

```
[100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200]
```

```
#Randomized SearchCV
#Number of trees in random forest
```

```
#Number of trees in random forest
#No of features to consider at every split
#max no of levels in tree
#min no of samples required to split a node
#min no of samples required at each leaf node


from sklearn.model_selection import RandomizedSearchCV
max_features=['auto','sqrt']
max_depth=[int(x) for x in np.linspace(5,30,num=6)]
min_samples_split=[2,5,10,15,100]
min_samples_leaf=[1,2,5,10]


from sklearn.model_selection import RandomizedSearchCV
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}
print(random_grid)
```

```
{'n_estimators': [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200], 'ma
```

```
rf=RandomForestRegressor()


rf_random=RandomizedSearchCV(estimator=rf,param_distributions=random_grid,scoring='neg_mean
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
Fitting 5 folds for each of 10 candidates, totalling 50 fits
[CV] END max_depth=10, max_features=sqrt, min_samples_leaf=5, min_samples_split=5, n_
[CV] END max_depth=10, max_features=sqrt, min_samples_leaf=5, min_samples_split=5, n_
[CV] END max_depth=10, max_features=sqrt, min_samples_leaf=5, min_samples_split=5, n_
[CV] END max_depth=10, max_features=sqrt, min_samples_leaf=5, min_samples_split=5, n_
[CV] END max_depth=10, max_features=sqrt, min_samples_leaf=5, min_samples_split=5, n_
[CV] END max_depth=15, max_features=sqrt, min_samples_leaf=2, min_samples_split=10, r
[CV] END max_depth=15, max_features=sqrt, min_samples_leaf=2, min_samples_split=10, r
[CV] END max_depth=15, max_features=sqrt, min_samples_leaf=2, min_samples_split=10, r
[CV] END max_depth=15, max_features=sqrt, min_samples_leaf=2, min_samples_split=10, r
[CV] END max_depth=15, max_features=sqrt, min_samples_leaf=2, min_samples_split=10, r
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=15, max_features=auto, min_samples_leaf=5, min_samples_split=100,
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=15, max_features=auto, min_samples_leaf=5, min_samples_split=100,
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=15, max_features=auto, min_samples_leaf=5, min_samples_split=100,
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
```

```
  warn(
[CV] END max_depth=15, max_features=auto, min_samples_leaf=5, min_samples_split=100,
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=15, max_features=auto, min_samples_leaf=5, min_samples_split=100,
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=15, max_features=auto, min_samples_leaf=5, min_samples_split=5, n_
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=15, max_features=auto, min_samples_leaf=5, min_samples_split=5, n_
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=15, max_features=auto, min_samples_leaf=5, min_samples_split=5, n_
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=15, max_features=auto, min_samples_leaf=5, min_samples_split=5, n_
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=15, max_features=auto, min_samples_leaf=5, min_samples_split=5, n_
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=20, max_features=auto, min_samples_leaf=10, min_samples_split=5, r
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=20, max_features=auto, min_samples_leaf=10, min_samples_split=5, r
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=20, max_features=auto, min_samples_leaf=10, min_samples_split=5, r
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
les_split=5, r
: FutureWarnir
  warn(
[CV] END max_depth=20, max_features=auto, min_samples_leaf=10, min_samples_split=5, r
[CV] END max_depth=25, max_features=sqrt, min_samples_leaf=1, min_samples_split=2, n_
[CV] END max_depth=25, max_features=sqrt, min_samples_leaf=1, min_samples_split=2, n_
[CV] END max_depth=25, max_features=sqrt, min_samples_leaf=1, min_samples_split=2, n_
[CV] END max_depth=25, max_features=sqrt, min_samples_leaf=1, min_samples_split=2, n_
[CV] END max_depth=25, max_features=sqrt, min_samples_leaf=1, min_samples_split=2, n_
[CV] END max_depth=5, max_features=sqrt, min_samples_leaf=10, min_samples_split=15, r
[CV] END max_depth=5, max_features=sqrt, min_samples_leaf=10, min_samples_split=15, r
[CV] END max_depth=5, max_features=sqrt, min_samples_leaf=10, min_samples_split=15, r
[CV] END max_depth=5, max_features=sqrt, min_samples_leaf=10, min_samples_split=15, r
[CV] END max_depth=5, max_features=sqrt, min_samples_leaf=10, min_samples_split=15, r
[CV] END max_depth=15, max_features=sqrt, min_samples_leaf=1, min_samples_split=15, r
[CV] END max_depth=15, max_features=sqrt, min_samples_leaf=1, min_samples_split=15, r
[CV] END max_depth=15, max_features=sqrt, min_samples_leaf=1, min_samples_split=15, r
[CV] END max_depth=15, max_features=sqrt, min_samples_leaf=1, min_samples_split=15, r
[CV] END max_depth=15, max_features=sqrt, min_samples_leaf=1, min_samples_split=15, r
[CV] END max_depth=5, max_features=sqrt, min_samples_leaf=2, min_samples_split=10, n_
[CV] END max_depth=5, max_features=sqrt, min_samples_leaf=2, min_samples_split=10, n_
[CV] END max_depth=5, max_features=sqrt, min_samples_leaf=2, min_samples_split=10, n_
[CV] END max_depth=5, max_features=sqrt, min_samples_leaf=2, min_samples_split=10, n_
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
[CV] END max_depth=5, max_features=sqrt, min_samples_leaf=2, min_samples_split=10, n_
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=20, max_features=auto, min_samples_leaf=1, min_samples_split=15, r
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=20, max_features=auto, min_samples_leaf=1, min_samples_split=15, r
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=20, max_features=auto, min_samples_leaf=1, min_samples_split=15, r
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=20, max_features=auto, min_samples_leaf=1, min_samples_split=15, r
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
[CV] END max_depth=20, max_features=auto, min_samples_leaf=1, min_samples_split=15, r
/usr/local/lib/python3.10/dist-packages/sklearn/ensemble/_forest.py:413: FutureWarnir
  warn(
```

▸ **RandomizedSearchCV**

▸ **estimator: RandomForestRegressor**

▸ RandomForestRegressor

```
predictions=rf_random.predict(x_test)
```

```
predictions
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

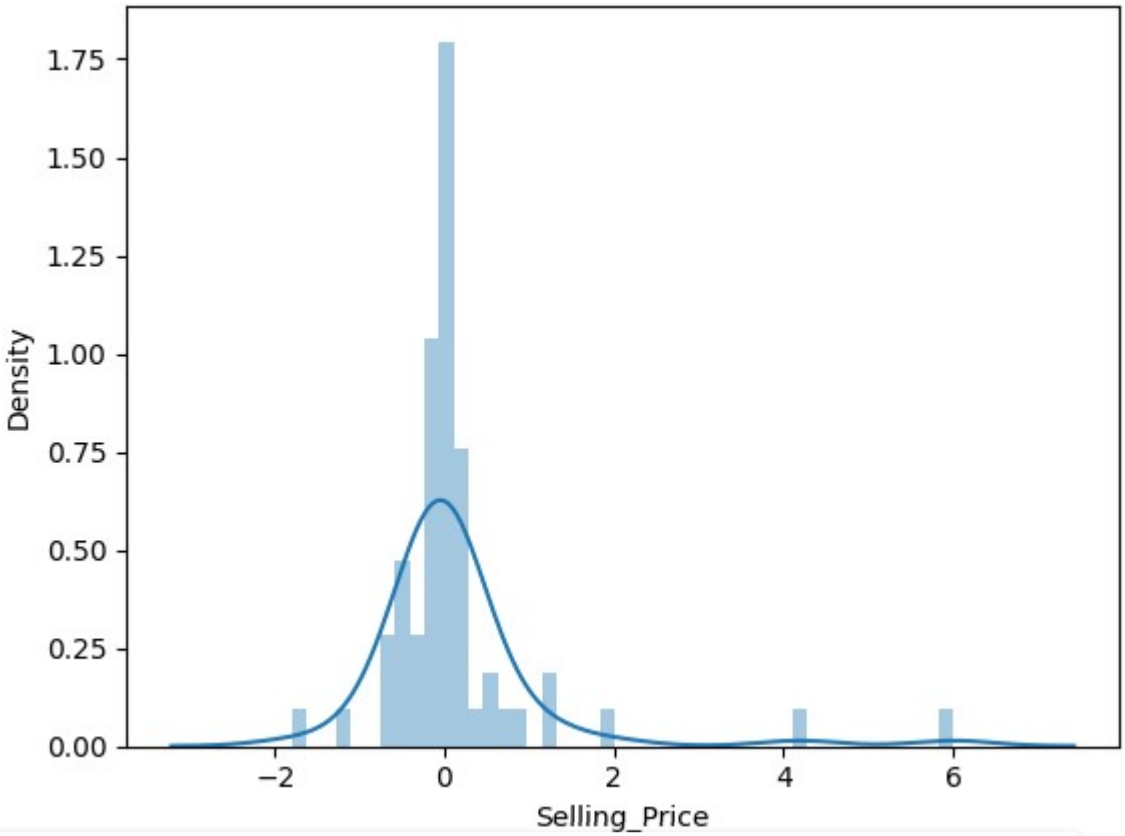
```
        4.41664763,  3.88350503,  7.02877722,  1.14930255,  4.40525744,
        6.17371586,  0.57725911,  3.37410033,  2.83497013,  0.2621329 ,
       10.44916056,  0.37179789,  2.85485748,  0.40121162,  0.30130336,
        0.52785394,  4.89398868,  0.42353471,  0.2621329 ,  0.25751815,
        4.3407383 ,  5.17977606,  1.16902903, 13.74241863,  1.16097731,
        8.53095945,  2.8644031 , 18.75977965,  0.55589027,  0.31035571,
       10.54056552,  0.38724278,  7.76784084,  5.76787314,  0.38394623,
       10.42196572,  5.54053519,  0.55978504,  0.48334168,  0.59776659,
        0.67310378,  1.14481057,  0.30711582,  2.91505144,  1.88371686,
        2.73517581,  0.63586022,  5.94522353,  7.05952898,  0.25751815,
        5.47114063,  5.74907968, 13.54528008,  0.31434599,  3.71314495,
        4.37271353])
```

```
sns.distplot(y_test-predictions)
plt.show()
```

```
<ipython-input-65-10d72eac0dde>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
```

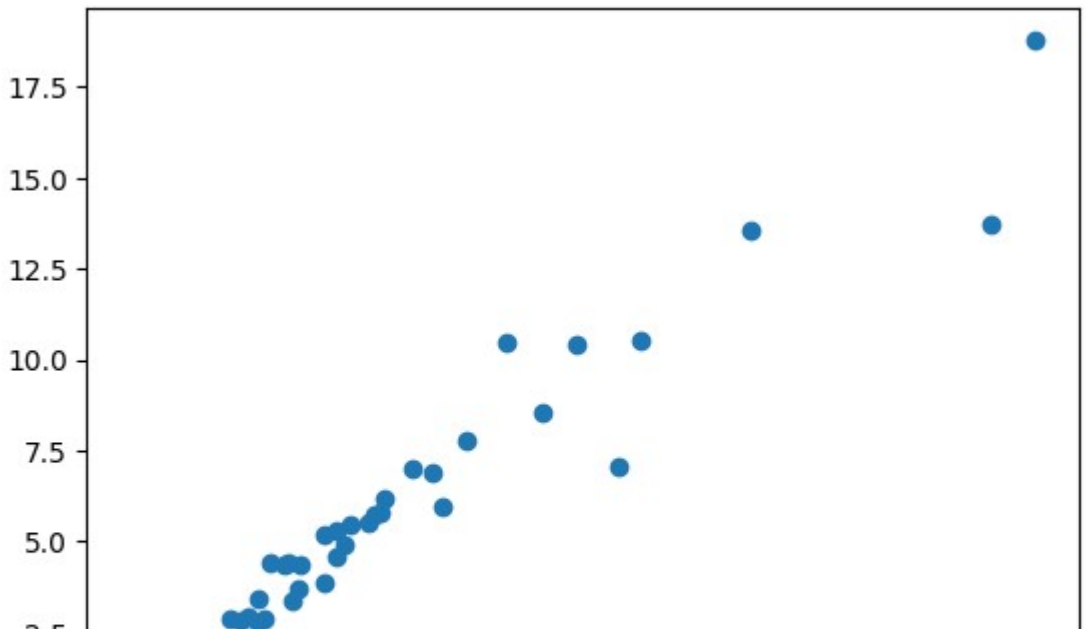
```
Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
```

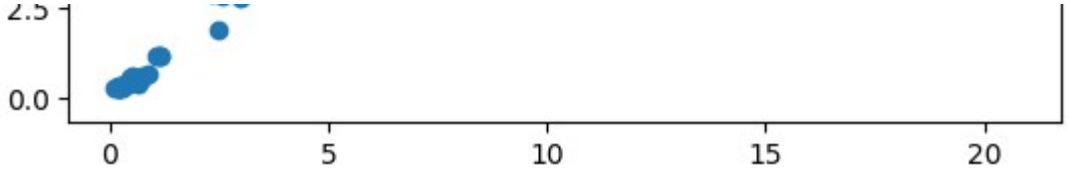
```
sns.distplot(y_test-predictions)
```



Automatic saving failed. This file was updated remotely or in another tab. Show diff

```
plt.scatter(y_test,predictions)
plt.show()
```






```
#Test model
x_new=np.array([[3.2,2000,0,5,1,0,0,1],[1.2,3000,0,1,0,1,0,1],[1.5,100000,0,9,0,1,1,0]])
predictions=rf_random.predict(x_new)
print("Prediction of price is:{}".format(predictions))
```

```
Prediction of price is:[1.94610381 1.20957438 1.05998333]
/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning: X does not
  warnings.warn(
```

Automatic saving failed. This file was updated remotely or in another tab. Show diff

Colab paid products - Cancel contracts here