House Price Prediction and EDA

December 4, 2022

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

     #!pip install dataprep

     from dataprep.eda import *

     import warnings
     warnings.filterwarnings('ignore')
```

```
[2]: train = pd.read_csv(r'C:\Users\adity\Downloads\DS Assignment\DS - Assignment␣
     ↪Part 1 data set.csv')
```

```
[3]: train.head()
```

```
[3]:    Transaction date  House Age  Distance from nearest Metro station (km)  \
     0           2012.917       32.0                                  84.87882
     1           2012.917       19.5                                 306.59470
     2           2013.583       13.3                                 561.98450
     3           2013.500       13.3                                 561.98450
     4           2012.833        5.0                                 390.56840

        Number of convenience stores  latitude  longitude  Number of bedrooms  \
     0                            10  24.98298  121.54024                   1
     1                             9  24.98034  121.53951                   2
     2                             5  24.98746  121.54391                   3
     3                             5  24.98746  121.54391                   2
     4                             5  24.97937  121.54245                   1

        House size (sqft)  House price of unit area
     0                575                      37.9
     1               1240                      42.2
     2               1060                      47.3
     3                875                      54.8
     4                491                      43.1
```

```
[4]: train.describe().T
```

```
[4]:                                             count          mean           std  \
     Transaction date                           414.0  2013.148971      0.281967
     House Age                                  414.0    17.712560     11.392485
     Distance from nearest Metro station (km)   414.0  1083.885689   1262.109595
     Number of convenience stores               414.0     4.094203      2.945562
     latitude                                   414.0    24.969030      0.012410
     longitude                                  414.0   121.533361      0.015347
     Number of bedrooms                         414.0     1.987923      0.818875
     House size (sqft)                          414.0   931.475845    348.910269
     House price of unit area                   414.0    37.980193     13.606488

                                                       min          25%          50%  \
     Transaction date                           2012.66700  2012.917000  2013.16700
     House Age                                     0.00000     9.025000    16.10000
     Distance from nearest Metro station (km)     23.38284   289.324800   492.23130
     Number of convenience stores                  0.00000     1.000000     4.00000
     latitude                                     24.93207    24.963000    24.97110
     longitude                                   121.47353   121.528085   121.53863
     Number of bedrooms                            1.00000     1.000000     2.00000
     House size (sqft)                           402.00000   548.000000   975.00000
     House price of unit area                      7.60000    27.700000    38.45000

                                                        75%          max
     Transaction date                           2013.417000  2013.58300
     House Age                                    28.150000    43.80000
     Distance from nearest Metro station (km)   1454.279000  6488.02100
     Number of convenience stores                 6.000000    10.00000
     latitude                                     24.977455    25.01459
     longitude                                   121.543305   121.56627
     Number of bedrooms                            3.000000     3.00000
     House size (sqft)                          1234.750000  1500.00000
     House price of unit area                     46.600000   117.50000
```

```
[5]: train.shape
```

```
[5]: (414, 9)
```

```
[6]: def check_df(dataframe, head=5):
         print("              Shape             ")
         print(dataframe.shape)

         print("              Types             ")
         print(dataframe.dtypes)

         print("              NA               ")
```

```
    print(dataframe.isnull().sum())


    print("           Quantiles           ")
    print(dataframe.quantile([0, 0.05, 0.50, 0.95, 0.99, 1]).T)
```

[7]: `check_df(train)`

```
          Shape
(414, 9)
          Types
Transaction date                          float64
House Age                                 float64
Distance from nearest Metro station (km)  float64
Number of convenience stores                int64
latitude                                  float64
longitude                                 float64
Number of bedrooms                          int64
House size (sqft)                           int64
House price of unit area                  float64
dtype: object
          NA
Transaction date                          0
House Age                                 0
Distance from nearest Metro station (km)  0
Number of convenience stores              0
latitude                                  0
longitude                                 0
Number of bedrooms                        0
House size (sqft)                         0
House price of unit area                  0
dtype: int64
          Quantiles
                                               0.00         0.05         0.50  \
Transaction date                          2012.66700  2012.667000  2013.16700
House Age                                    0.00000     1.100000    16.10000
Distance from nearest Metro station (km)    23.38284    90.456060   492.23130
Number of convenience stores                 0.00000     0.000000     4.00000
latitude                                    24.93207    24.945759    24.97110
longitude                                  121.47353   121.503420   121.53863
Number of bedrooms                           1.00000     1.000000     2.00000
House size (sqft)                          402.00000   434.950000   975.00000
House price of unit area                     7.60000    16.490000    38.45000


                                                0.95         0.99         1.00
Transaction date                          2013.583000  2013.583000  2013.58300
House Age                                    37.735000    41.248000    43.80000
Distance from nearest Metro station (km)   4082.015000  5394.220430  6488.02100
Number of convenience stores                  9.000000    10.000000    10.00000
```

```
latitude                                        24.985704     24.991734     25.01459
longitude                                      121.549053    121.559012    121.56627
Number of bedrooms                               3.000000      3.000000      3.00000
House size (sqft)                             1438.350000   1493.610000   1500.00000
House price of unit area                        59.175000     70.883000    117.50000
```
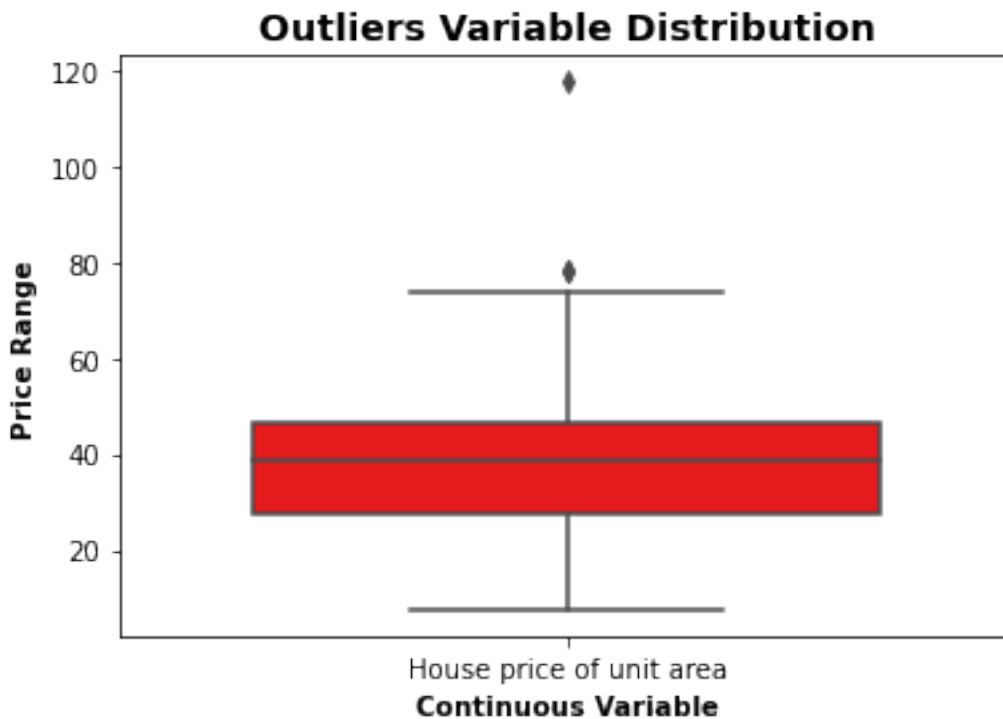
[8]: `train.duplicated().sum()`

[8]: 0

[9]:
```python
outliers = ['House price of unit area']
plt.rcParams['figure.figsize'] = [8,8]
sns.boxplot(data=train[outliers], orient="v", palette="Set1" ,whis=1.
 ↪5,saturation=1, width=0.7)
plt.title("Outliers Variable Distribution", fontsize = 14, fontweight = 'bold')
plt.ylabel("Price Range", fontweight = 'bold')
plt.xlabel("Continuous Variable", fontweight = 'bold')
train.shape
```

[9]: (414, 9)

[10]:
```python
# Segregation of Numerical and Categorical Variables/Columns
cat_col = train.select_dtypes(include=['object']).columns
num_col = train.select_dtypes(exclude=['object']).columns
```

```
[11]: cat_col
```

```
[11]: Index([], dtype='object')
```

```
[12]: num_col
```

```
[12]: Index(['Transaction date', 'House Age',
            'Distance from nearest Metro station (km)',
            'Number of convenience stores', 'latitude', 'longitude',
            'Number of bedrooms', 'House size (sqft)', 'House price of unit area'],
           dtype='object')
```

```
[13]: plt.figure(figsize=(10,10))
      plt.title("House Price of Unit Area Distribution Plot")
      sns.distplot(train["House price of unit area"])
```

```
[13]: <AxesSubplot:title={'center':'House Price of Unit Area Distribution Plot'},
      xlabel='House price of unit area', ylabel='Density'>
```

House Price of Unit Area Distribution Plot
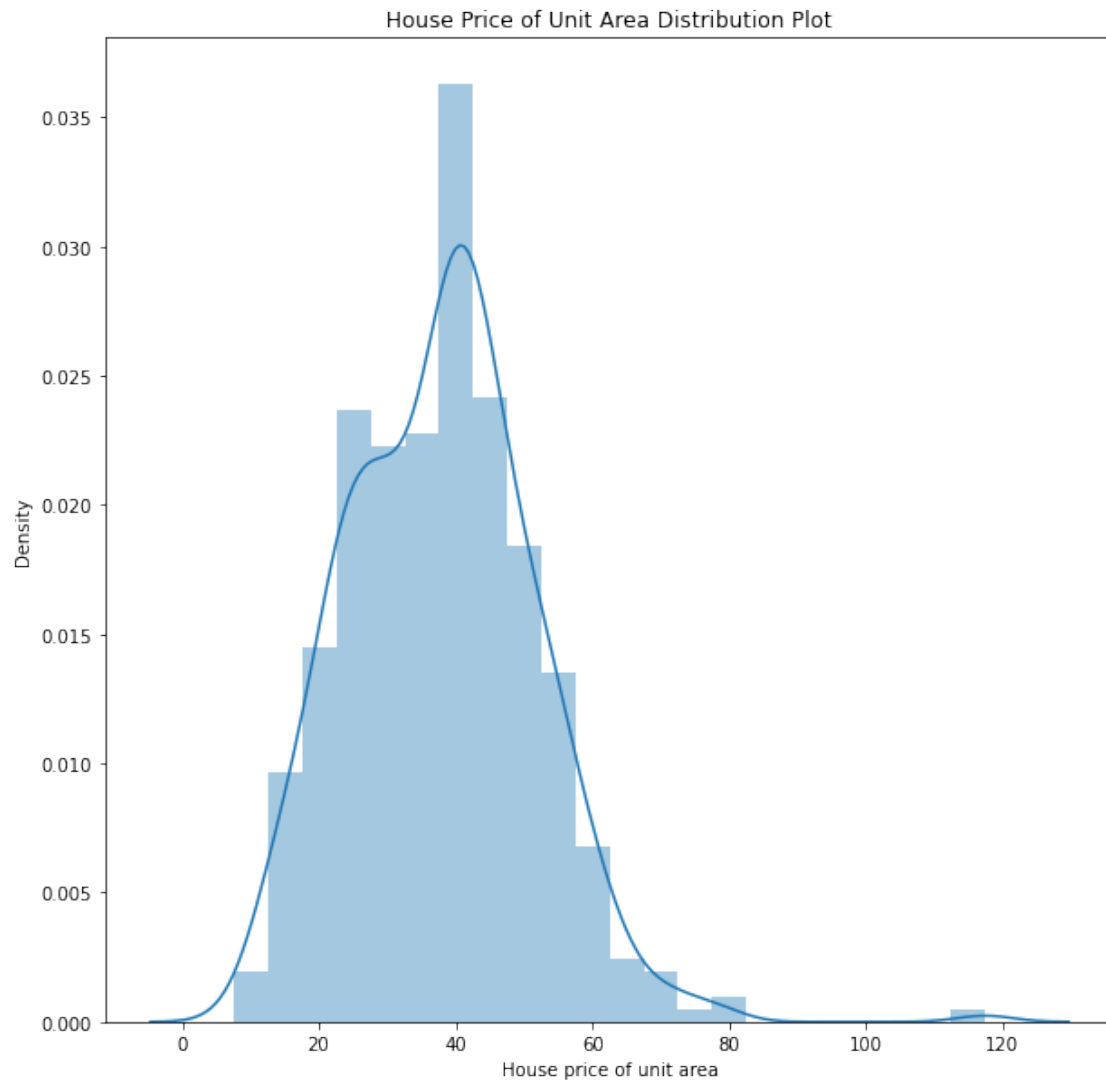
```
[14]: plot(train)
```

      0%| | 0/538 [00:00<?, ?it/s]

```
[14]: <dataprep.eda.container.Container at 0x23b87112430>
```

```
[15]: plot_correlation(train, "House price of unit area")
```

      0%| | 0/203 [00:00<?, ?it/s]

```
[15]: <dataprep.eda.container.Container at 0x23b875ca400>
```

```
[16]: plot_correlation(train)
```

```
[16]: <dataprep.eda.container.Container at 0x23b8715c7f0>
```

```
[17]: plot_correlation(train[["House price of unit area","Number of convenience␣
       ↪stores","latitude","longitude","House size (sqft)"]])
```

[17]: <dataprep.eda.container.Container at 0x23b875f1f10>

```
[18]: #UNIVARIATE ANALYSIS __>
      plot(train, "House price of unit area")
```

```
 0%|           | 0/122 [00:00<?, ?it/s]
```

[18]: <dataprep.eda.container.Container at 0x23b87532e50>

Pivotal Features

```
[19]: plot_correlation(train, "Number of convenience stores", "House price of unit␣
       ↪area")
```

```
 0%|           | 0/12 [00:00<?, ?it/s]
```

[19]: <dataprep.eda.container.Container at 0x23b87286cd0>

```
[20]: plot(train, "Number of convenience stores", "House price of unit area")
```

```
 0%|           | 0/9 [00:00<?, ?it/s]
```

[20]: <dataprep.eda.container.Container at 0x23b885ef6d0>

```
[21]: plot(train,"latitude","House price of unit area")
```

```
 0%|           | 0/9 [00:00<?, ?it/s]
```

[21]: <dataprep.eda.container.Container at 0x23b9071b040>

```
[22]: plot(train,"longitude","House price of unit area")
```

```
 0%|           | 0/9 [00:00<?, ?it/s]
```

[22]: <dataprep.eda.container.Container at 0x23b88635790>

```
[23]: #Prediction of House Unit Price by Pycaret module
```

```
[24]: !pip install pycaret
      from pycaret.utils import enable_colab
      from pycaret.classification import *
```

```
Requirement already satisfied: pycaret in
c:\users\adity\appdata\roaming\python\python38\site-packages (2.3.10)
Requirement already satisfied: scikit-plot in
c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret)
(0.3.7)
Requirement already satisfied: lightgbm>=2.3.1 in
c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret)
```

(3.3.3)
Requirement already satisfied: textblob in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (0.17.1)
Requirement already satisfied: IPython in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (8.4.0)
Requirement already satisfied: ipywidgets in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (7.7.2)
Requirement already satisfied: nltk in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (3.7)
Requirement already satisfied: umap-learn in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (0.5.3)
Requirement already satisfied: imbalanced-learn==0.7.0 in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (0.7.0)
Requirement already satisfied: joblib in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (1.1.0)
Requirement already satisfied: pyLDAvis in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (3.2.2)
Requirement already satisfied: scipy<=1.5.4 in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (1.5.4)
Requirement already satisfied: pandas in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (1.4.0)
Requirement already satisfied: spacy<2.4.0 in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (2.3.8)
Requirement already satisfied: plotly>=4.4.1 in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (4.8.1)
Requirement already satisfied: gensim<4.0.0 in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (3.8.3)
Requirement already satisfied: kmodes>=0.10.1 in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (0.12.2)
Requirement already satisfied: cufflinks>=0.17.0 in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (0.17.3)
Requirement already satisfied: mlflow in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (2.0.1)
Requirement already satisfied: yellowbrick>=1.0.1 in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (1.3.post1)
Requirement already satisfied: scikit-learn==0.23.2 in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (0.23.2)
Requirement already satisfied: mlxtend>=0.17.0 in

c:\users\adity\anaconda3\lib\site-packages (from pycaret) (0.19.0)
Requirement already satisfied: wordcloud in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (1.8.2.2)
Requirement already satisfied: numba<0.55 in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (0.54.1)
Requirement already satisfied: matplotlib in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (3.5.3)
Requirement already satisfied: pyod in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (1.0.1)
Requirement already satisfied: Boruta in c:\users\adity\appdata\roaming\python\python38\site-packages (from pycaret) (0.3)
Requirement already satisfied: seaborn in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (0.11.2)
Requirement already satisfied: pandas-profiling>=2.8.0 in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (2.13.0)
Requirement already satisfied: pyyaml<6.0.0 in c:\users\adity\anaconda3\lib\site-packages (from pycaret) (5.4.1)
Requirement already satisfied: numpy>=1.13.3 in c:\users\adity\anaconda3\lib\site-packages (from imbalanced-learn==0.7.0->pycaret) (1.19.5)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\adity\anaconda3\lib\site-packages (from scikit-learn==0.23.2->pycaret) (3.1.0)
Requirement already satisfied: setuptools>=34.4.1 in c:\users\adity\anaconda3\lib\site-packages (from cufflinks>=0.17.0->pycaret) (61.2.0)
Requirement already satisfied: six>=1.9.0 in c:\users\adity\anaconda3\lib\site-packages (from cufflinks>=0.17.0->pycaret) (1.16.0)
Requirement already satisfied: colorlover>=0.2.1 in c:\users\adity\anaconda3\lib\site-packages (from cufflinks>=0.17.0->pycaret) (0.3.0)
Requirement already satisfied: Cython==0.29.14 in c:\users\adity\anaconda3\lib\site-packages (from gensim<4.0.0->pycaret) (0.29.14)
Requirement already satisfied: smart-open>=1.8.1 in c:\users\adity\anaconda3\lib\site-packages (from gensim<4.0.0->pycaret) (5.2.1)
Requirement already satisfied: jedi>=0.16 in c:\users\adity\anaconda3\lib\site-packages (from IPython->pycaret) (0.18.1)
Requirement already satisfied: traitlets>=5 in c:\users\adity\anaconda3\lib\site-packages (from IPython->pycaret) (5.1.1)
Requirement already satisfied: pickleshare in c:\users\adity\anaconda3\lib\site-packages (from IPython->pycaret) (0.7.5)
Requirement already satisfied: prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0 in c:\users\adity\anaconda3\lib\site-packages (from IPython->pycaret) (3.0.20)
Requirement already satisfied: backcall in c:\users\adity\anaconda3\lib\site-packages (from IPython->pycaret) (0.2.0)

Requirement already satisfied: colorama in c:\users\adity\anaconda3\lib\site-packages (from IPython->pycaret) (0.4.4)
Requirement already satisfied: matplotlib-inline in c:\users\adity\anaconda3\lib\site-packages (from IPython->pycaret) (0.1.2)
Requirement already satisfied: decorator in c:\users\adity\anaconda3\lib\site-packages (from IPython->pycaret) (5.1.1)
Requirement already satisfied: pygments>=2.4.0 in c:\users\adity\anaconda3\lib\site-packages (from IPython->pycaret) (2.11.2)
Requirement already satisfied: stack-data in c:\users\adity\anaconda3\lib\site-packages (from IPython->pycaret) (0.2.0)
Requirement already satisfied: widgetsnbextension~=3.6.0 in c:\users\adity\anaconda3\lib\site-packages (from ipywidgets->pycaret) (3.6.1)
Requirement already satisfied: ipykernel>=4.5.1 in c:\users\adity\anaconda3\lib\site-packages (from ipywidgets->pycaret) (6.9.1)
Requirement already satisfied: jupyterlab-widgets<3,>=1.0.0 in c:\users\adity\anaconda3\lib\site-packages (from ipywidgets->pycaret) (1.1.1)
Requirement already satisfied: ipython-genutils~=0.2.0 in c:\users\adity\anaconda3\lib\site-packages (from ipywidgets->pycaret) (0.2.0)
Requirement already satisfied: wheel in c:\users\adity\anaconda3\lib\site-packages (from lightgbm>=2.3.1->pycaret) (0.37.1)
Requirement already satisfied: pyparsing>=2.2.1 in c:\users\adity\anaconda3\lib\site-packages (from matplotlib->pycaret) (3.0.4)
Requirement already satisfied: cycler>=0.10 in c:\users\adity\anaconda3\lib\site-packages (from matplotlib->pycaret) (0.11.0)
Requirement already satisfied: python-dateutil>=2.7 in c:\users\adity\anaconda3\lib\site-packages (from matplotlib->pycaret) (2.8.2)
Requirement already satisfied: pillow>=6.2.0 in c:\users\adity\anaconda3\lib\site-packages (from matplotlib->pycaret) (9.2.0)
Requirement already satisfied: fonttools>=4.22.0 in c:\users\adity\anaconda3\lib\site-packages (from matplotlib->pycaret) (4.34.4)
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\adity\anaconda3\lib\site-packages (from matplotlib->pycaret) (1.4.4)
Requirement already satisfied: packaging>=20.0 in c:\users\adity\anaconda3\lib\site-packages (from matplotlib->pycaret) (21.3)
Requirement already satisfied: llvmlite<0.38,>=0.37.0rc1 in c:\users\adity\appdata\roaming\python\python38\site-packages (from numba<0.55->pycaret) (0.37.0)
Requirement already satisfied: pytz>=2020.1 in c:\users\adity\anaconda3\lib\site-packages (from pandas->pycaret) (2022.2.1)
Requirement already satisfied: missingno>=0.4.2 in c:\users\adity\anaconda3\lib\site-packages (from pandas-profiling>=2.8.0->pycaret) (0.5.1)
Requirement already satisfied: confuse>=1.0.0 in c:\users\adity\anaconda3\lib\site-packages (from pandas-profiling>=2.8.0->pycaret) (2.0.0)
Requirement already satisfied: tangled-up-in-unicode>=0.0.6 in c:\users\adity\anaconda3\lib\site-packages (from pandas-profiling>=2.8.0->pycaret) (0.2.0)

```
Requirement already satisfied: tqdm>=4.48.2 in
c:\users\adity\anaconda3\lib\site-packages (from pandas-
profiling>=2.8.0->pycaret) (4.64.1)
Requirement already satisfied: htmlmin>=0.1.12 in
c:\users\adity\anaconda3\lib\site-packages (from pandas-
profiling>=2.8.0->pycaret) (0.1.12)
Requirement already satisfied: attrs>=19.3.0 in
c:\users\adity\anaconda3\lib\site-packages (from pandas-
profiling>=2.8.0->pycaret) (21.4.0)
Requirement already satisfied: jinja2>=2.11.1 in
c:\users\adity\anaconda3\lib\site-packages (from pandas-
profiling>=2.8.0->pycaret) (3.0.3)
Requirement already satisfied: phik>=0.11.1 in
c:\users\adity\anaconda3\lib\site-packages (from pandas-
profiling>=2.8.0->pycaret) (0.12.2)
Requirement already satisfied: requests>=2.24.0 in
c:\users\adity\anaconda3\lib\site-packages (from pandas-
profiling>=2.8.0->pycaret) (2.28.1)
Requirement already satisfied: visions[type_image_path]==0.7.1 in
c:\users\adity\anaconda3\lib\site-packages (from pandas-
profiling>=2.8.0->pycaret) (0.7.1)
Requirement already satisfied: networkx>=2.4 in
c:\users\adity\anaconda3\lib\site-packages (from
visions[type_image_path]==0.7.1->pandas-profiling>=2.8.0->pycaret) (2.8.5)
Requirement already satisfied: bottleneck in c:\users\adity\anaconda3\lib\site-
packages (from visions[type_image_path]==0.7.1->pandas-
profiling>=2.8.0->pycaret) (1.3.5)
Requirement already satisfied: multimethod==1.4 in
c:\users\adity\anaconda3\lib\site-packages (from
visions[type_image_path]==0.7.1->pandas-profiling>=2.8.0->pycaret) (1.4)
Requirement already satisfied: imagehash in c:\users\adity\anaconda3\lib\site-
packages (from visions[type_image_path]==0.7.1->pandas-
profiling>=2.8.0->pycaret) (4.2.1)
Requirement already satisfied: retrying>=1.3.3 in
c:\users\adity\anaconda3\lib\site-packages (from plotly>=4.4.1->pycaret) (1.3.3)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in
c:\users\adity\anaconda3\lib\site-packages (from spacy<2.4.0->pycaret) (1.0.7)
Requirement already satisfied: blis<0.8.0,>=0.4.0 in
c:\users\adity\anaconda3\lib\site-packages (from spacy<2.4.0->pycaret) (0.7.7)
Requirement already satisfied: thinc<7.5.0,>=7.4.1 in
c:\users\adity\appdata\roaming\python\python38\site-packages (from
spacy<2.4.0->pycaret) (7.4.6)
Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in
c:\users\adity\anaconda3\lib\site-packages (from spacy<2.4.0->pycaret) (1.0.2)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in
c:\users\adity\anaconda3\lib\site-packages (from spacy<2.4.0->pycaret) (3.0.6)
Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in
c:\users\adity\anaconda3\lib\site-packages (from spacy<2.4.0->pycaret) (0.9.1)
```

Requirement already satisfied: srsly<1.1.0,>=1.0.2 in
c:\users\adity\anaconda3\lib\site-packages (from spacy<2.4.0->pycaret) (1.0.6)
Requirement already satisfied: plac<1.2.0,>=0.9.6 in
c:\users\adity\anaconda3\lib\site-packages (from spacy<2.4.0->pycaret) (1.1.3)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in
c:\users\adity\anaconda3\lib\site-packages (from spacy<2.4.0->pycaret) (2.0.6)
Requirement already satisfied: sqlparse<1,>=0.4.0 in
c:\users\adity\anaconda3\lib\site-packages (from mlflow->pycaret) (0.4.2)
Requirement already satisfied: gitpython<4,>=2.1.0 in
c:\users\adity\appdata\roaming\python\python38\site-packages (from
mlflow->pycaret) (3.1.29)
Requirement already satisfied: alembic<2 in
c:\users\adity\appdata\roaming\python\python38\site-packages (from
mlflow->pycaret) (1.8.1)
Requirement already satisfied: markdown<4,>=3.3 in
c:\users\adity\anaconda3\lib\site-packages (from mlflow->pycaret) (3.3.6)
Requirement already satisfied: Flask<3 in c:\users\adity\anaconda3\lib\site-
packages (from mlflow->pycaret) (2.2.2)
Requirement already satisfied: docker<7,>=4.0.0 in
c:\users\adity\appdata\roaming\python\python38\site-packages (from
mlflow->pycaret) (6.0.1)
Requirement already satisfied: pyarrow<11,>=4.0.0 in
c:\users\adity\appdata\roaming\python\python38\site-packages (from
mlflow->pycaret) (10.0.1)
Requirement already satisfied: databricks-cli<1,>=0.8.7 in
c:\users\adity\appdata\roaming\python\python38\site-packages (from
mlflow->pycaret) (0.17.3)
Requirement already satisfied: sqlalchemy<2,>=1.4.0 in
c:\users\adity\anaconda3\lib\site-packages (from mlflow->pycaret) (1.4.44)
Requirement already satisfied: entrypoints<1 in
c:\users\adity\anaconda3\lib\site-packages (from mlflow->pycaret) (0.4)
Requirement already satisfied: cloudpickle<3 in
c:\users\adity\anaconda3\lib\site-packages (from mlflow->pycaret) (2.2.0)
Requirement already satisfied: querystring-parser<2 in
c:\users\adity\anaconda3\lib\site-packages (from mlflow->pycaret) (1.2.4)
Requirement already satisfied: protobuf<5,>=3.12.0 in
c:\users\adity\anaconda3\lib\site-packages (from mlflow->pycaret) (3.19.1)
Requirement already satisfied: click<9,>=7.0 in
c:\users\adity\anaconda3\lib\site-packages (from mlflow->pycaret) (8.0.3)
Requirement already satisfied: shap<1,>=0.40 in
c:\users\adity\appdata\roaming\python\python38\site-packages (from
mlflow->pycaret) (0.41.0)
Requirement already satisfied: importlib-metadata!=4.7.0,<6,>=3.7.0 in
c:\users\adity\anaconda3\lib\site-packages (from mlflow->pycaret) (4.12.0)
Requirement already satisfied: waitress<3 in c:\users\adity\anaconda3\lib\site-
packages (from mlflow->pycaret) (2.1.2)
Requirement already satisfied: regex>=2021.8.3 in
c:\users\adity\anaconda3\lib\site-packages (from nltk->pycaret) (2021.11.10)

```
Requirement already satisfied: numexpr in c:\users\adity\anaconda3\lib\site-
packages (from pyLDAvis->pycaret) (2.8.3)
Requirement already satisfied: funcy in c:\users\adity\anaconda3\lib\site-
packages (from pyLDAvis->pycaret) (1.17)
Requirement already satisfied: future in c:\users\adity\anaconda3\lib\site-
packages (from pyLDAvis->pycaret) (0.18.2)
Requirement already satisfied: statsmodels in
c:\users\adity\appdata\roaming\python\python38\site-packages (from
pyod->pycaret) (0.13.5)
Requirement already satisfied: pynndescent>=0.5 in
c:\users\adity\appdata\roaming\python\python38\site-packages (from umap-
learn->pycaret) (0.5.8)
Requirement already satisfied: importlib-resources in
c:\users\adity\anaconda3\lib\site-packages (from alembic<2->mlflow->pycaret)
(5.2.0)
Requirement already satisfied: Mako in c:\users\adity\anaconda3\lib\site-
packages (from alembic<2->mlflow->pycaret) (1.2.4)
Requirement already satisfied: pyjwt>=1.7.0 in
c:\users\adity\anaconda3\lib\site-packages (from databricks-
cli<1,>=0.8.7->mlflow->pycaret) (2.3.0)
Requirement already satisfied: oauthlib>=3.1.0 in
c:\users\adity\anaconda3\lib\site-packages (from databricks-
cli<1,>=0.8.7->mlflow->pycaret) (3.2.0)
Requirement already satisfied: tabulate>=0.7.7 in
c:\users\adity\anaconda3\lib\site-packages (from databricks-
cli<1,>=0.8.7->mlflow->pycaret) (0.9.0)
Requirement already satisfied: websocket-client>=0.32.0 in
c:\users\adity\anaconda3\lib\site-packages (from
docker<7,>=4.0.0->mlflow->pycaret) (0.58.0)
Requirement already satisfied: urllib3>=1.26.0 in
c:\users\adity\anaconda3\lib\site-packages (from
docker<7,>=4.0.0->mlflow->pycaret) (1.26.10)
Requirement already satisfied: pywin32>=304 in
c:\users\adity\anaconda3\lib\site-packages (from
docker<7,>=4.0.0->mlflow->pycaret) (305)
Requirement already satisfied: Werkzeug>=2.2.2 in
c:\users\adity\anaconda3\lib\site-packages (from Flask<3->mlflow->pycaret)
(2.2.2)
Requirement already satisfied: itsdangerous>=2.0 in
c:\users\adity\anaconda3\lib\site-packages (from Flask<3->mlflow->pycaret)
(2.0.1)
Requirement already satisfied: gitdb<5,>=4.0.1 in
c:\users\adity\appdata\roaming\python\python38\site-packages (from
gitpython<4,>=2.1.0->mlflow->pycaret) (4.0.9)
Requirement already satisfied: zipp>=0.5 in c:\users\adity\anaconda3\lib\site-
packages (from importlib-metadata!=4.7.0,<6,>=3.7.0->mlflow->pycaret) (3.8.1)
Requirement already satisfied: tornado<7.0,>=4.2 in
c:\users\adity\anaconda3\lib\site-packages (from
```

```
ipykernel>=4.5.1->ipywidgets->pycaret) (6.2)
Requirement already satisfied: nest-asyncio in
c:\users\adity\anaconda3\lib\site-packages (from
ipykernel>=4.5.1->ipywidgets->pycaret) (1.5.5)
Requirement already satisfied: debugpy<2.0,>=1.0.0 in
c:\users\adity\anaconda3\lib\site-packages (from
ipykernel>=4.5.1->ipywidgets->pycaret) (1.5.1)
Requirement already satisfied: jupyter-client<8.0 in
c:\users\adity\anaconda3\lib\site-packages (from
ipykernel>=4.5.1->ipywidgets->pycaret) (7.2.2)
Requirement already satisfied: parso<0.9.0,>=0.8.0 in
c:\users\adity\anaconda3\lib\site-packages (from jedi>=0.16->IPython->pycaret)
(0.8.3)
Requirement already satisfied: MarkupSafe>=2.0 in
c:\users\adity\anaconda3\lib\site-packages (from jinja2>=2.11.1->pandas-
profiling>=2.8.0->pycaret) (2.1.1)
Requirement already satisfied: wcwidth in c:\users\adity\anaconda3\lib\site-
packages (from prompt-toolkit!=3.0.0,!=3.0.1,<3.1.0,>=2.0.0->IPython->pycaret)
(0.2.5)
Requirement already satisfied: charset-normalizer<3,>=2 in
c:\users\adity\anaconda3\lib\site-packages (from requests>=2.24.0->pandas-
profiling>=2.8.0->pycaret) (2.0.11)
Requirement already satisfied: certifi>=2017.4.17 in
c:\users\adity\anaconda3\lib\site-packages (from requests>=2.24.0->pandas-
profiling>=2.8.0->pycaret) (2021.10.8)
Requirement already satisfied: idna<4,>=2.5 in
c:\users\adity\anaconda3\lib\site-packages (from requests>=2.24.0->pandas-
profiling>=2.8.0->pycaret) (3.3)
Requirement already satisfied: slicer==0.0.7 in
c:\users\adity\anaconda3\lib\site-packages (from shap<1,>=0.40->mlflow->pycaret)
(0.0.7)
Requirement already satisfied: greenlet!=0.4.17 in
c:\users\adity\anaconda3\lib\site-packages (from
sqlalchemy<2,>=1.4.0->mlflow->pycaret) (1.1.3.post0)
Requirement already satisfied: notebook>=4.4.1 in
c:\users\adity\anaconda3\lib\site-packages (from
widgetsnbextension~=3.6.0->ipywidgets->pycaret) (6.4.11)
Requirement already satisfied: asttokens in c:\users\adity\anaconda3\lib\site-
packages (from stack-data->IPython->pycaret) (2.0.5)
Requirement already satisfied: executing in c:\users\adity\anaconda3\lib\site-
packages (from stack-data->IPython->pycaret) (0.8.3)
Requirement already satisfied: pure-eval in c:\users\adity\anaconda3\lib\site-
packages (from stack-data->IPython->pycaret) (0.2.2)
Requirement already satisfied: patsy>=0.5.2 in
c:\users\adity\appdata\roaming\python\python38\site-packages (from
statsmodels->pyod->pycaret) (0.5.3)
Requirement already satisfied: smmap<6,>=3.0.1 in
c:\users\adity\anaconda3\lib\site-packages (from
```

gitdb<5,>=4.0.1->gitpython<4,>=2.1.0->mlflow->pycaret) (5.0.0)
Requirement already satisfied: jupyter-core>=4.9.2 in
c:\users\adity\anaconda3\lib\site-packages (from jupyter-
client<8.0->ipykernel>=4.5.1->ipywidgets->pycaret) (4.10.0)
Requirement already satisfied: pyzmq>=22.3 in c:\users\adity\anaconda3\lib\site-
packages (from jupyter-client<8.0->ipykernel>=4.5.1->ipywidgets->pycaret)
(23.2.0)
Requirement already satisfied: argon2-cffi in c:\users\adity\anaconda3\lib\site-
packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(21.3.0)
Requirement already satisfied: nbformat in c:\users\adity\anaconda3\lib\site-
packages (from notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(5.3.0)
Requirement already satisfied: terminado>=0.8.3 in
c:\users\adity\anaconda3\lib\site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret) (0.13.1)
Requirement already satisfied: prometheus-client in
c:\users\adity\anaconda3\lib\site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret) (0.13.1)
Requirement already satisfied: Send2Trash>=1.8.0 in
c:\users\adity\anaconda3\lib\site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret) (1.8.0)
Requirement already satisfied: nbconvert>=5 in
c:\users\adity\anaconda3\lib\site-packages (from
notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret) (6.4.4)
Requirement already satisfied: PyWavelets in c:\users\adity\anaconda3\lib\site-
packages (from imagehash->visions[type_image_path]==0.7.1->pandas-
profiling>=2.8.0->pycaret) (1.3.0)
Requirement already satisfied: bleach in c:\users\adity\anaconda3\lib\site-
packages (from
nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(4.1.0)
Requirement already satisfied: pandocfilters>=1.4.1 in
c:\users\adity\anaconda3\lib\site-packages (from
nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(1.5.0)
Requirement already satisfied: testpath in c:\users\adity\anaconda3\lib\site-
packages (from
nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(0.6.0)
Requirement already satisfied: jupyterlab-pygments in
c:\users\adity\anaconda3\lib\site-packages (from
nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(0.1.2)
Requirement already satisfied: beautifulsoup4 in
c:\users\adity\anaconda3\lib\site-packages (from
nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(4.11.1)

Requirement already satisfied: mistune<2,>=0.8.1 in
c:\users\adity\anaconda3\lib\site-packages (from
nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(0.8.4)
Requirement already satisfied: nbclient<0.6.0,>=0.5.0 in
c:\users\adity\anaconda3\lib\site-packages (from
nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(0.5.13)
Requirement already satisfied: defusedxml in c:\users\adity\anaconda3\lib\site-
packages (from
nbconvert>=5->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(0.7.1)
Requirement already satisfied: jsonschema>=2.6 in
c:\users\adity\anaconda3\lib\site-packages (from
nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(4.4.0)
Requirement already satisfied: fastjsonschema in
c:\users\adity\anaconda3\lib\site-packages (from
nbformat->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(2.15.1)
Requirement already satisfied: pywinpty>=1.1.0 in
c:\users\adity\anaconda3\lib\site-packages (from terminado>=0.8.3->notebook>=4.4
.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret) (2.0.2)
Requirement already satisfied: argon2-cffi-bindings in
c:\users\adity\anaconda3\lib\site-packages (from
argon2-cffi->notebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret)
(21.2.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in
c:\users\adity\anaconda3\lib\site-packages (from jsonschema>=2.6->nbformat->note
book>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret) (0.18.0)
Requirement already satisfied: cffi>=1.0.1 in c:\users\adity\anaconda3\lib\site-
packages (from argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->widgetsnbexte
nsion~=3.6.0->ipywidgets->pycaret) (1.15.0)
Requirement already satisfied: soupsieve>1.2 in
c:\users\adity\anaconda3\lib\site-packages (from beautifulsoup4->nbconvert>=5->n
otebook>=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret) (2.3.2.post1)
Requirement already satisfied: webencodings in
c:\users\adity\anaconda3\lib\site-packages (from bleach->nbconvert>=5->notebook>
=4.4.1->widgetsnbextension~=3.6.0->ipywidgets->pycaret) (0.5.1)
Requirement already satisfied: pycparser in c:\users\adity\anaconda3\lib\site-
packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi->notebook>=4.4.1->
widgetsnbextension~=3.6.0->ipywidgets->pycaret) (2.21)

[27]:
```python
#!pip install --user numba --upgrade
from pycaret.regression import *
```

```
[74]: #!pip uninstall scikit-learn
      !python -m pip uninstall sklearn --yes
      #!python -m pip uninstall numba --yes
```

```
Found existing installation: sklearn 0.0
Uninstalling sklearn-0.0:
  Successfully uninstalled sklearn-0.0
```

```
[76]: !pip install --user -U scikit-learn==0.23.2
```

```
Collecting scikit-learn==0.23.2
  Using cached scikit_learn-0.23.2-cp38-cp38-win_amd64.whl (6.8 MB)
Requirement already satisfied: joblib>=0.11 in
c:\users\adity\anaconda3\lib\site-packages (from scikit-learn==0.23.2) (1.1.0)
Requirement already satisfied: numpy>=1.13.3 in
c:\users\adity\anaconda3\lib\site-packages (from scikit-learn==0.23.2) (1.19.5)
Requirement already satisfied: threadpoolctl>=2.0.0 in
c:\users\adity\anaconda3\lib\site-packages (from scikit-learn==0.23.2) (3.1.0)
Requirement already satisfied: scipy>=0.19.1 in
c:\users\adity\anaconda3\lib\site-packages (from scikit-learn==0.23.2) (1.5.4)
Installing collected packages: scikit-learn
Successfully installed scikit-learn-0.23.2
```

```
[28]: HousePrice_data = setup(data = train,  target = 'House price of unit area',␣
      ↪feature_selection=True,␣
      ↪normalize=True,pca=True,remove_multicollinearity=True, session_id = 123)
```

```
<pandas.io.formats.style.Styler at 0x23b984fe370>
```

```
create_model_container: 0
master_model_container: 0
display_container: 1
Pipeline(memory=None,
         steps=[('dtypes',
                 DataTypes_Auto_infer(categorical_features=[],
                                      display_types=True, features_todrop=[],
                                      id_columns=[], ml_usecase='regression',
                                      numerical_features=[],
                                      target='House price of unit area',
                                      time_features=[])),
                ('imputer',
                 Simple_Imputer(categorical_strategy='not_available',
                                fill_value_categorical=None,
                                fill_value_numerical=None,
                                num…
                Fix_multicollinearity(correlation_with_target_preference=None,
                                      correlation_with_target_threshold=0.0,
                                      target_variable='House price of unit '
                                                      'area',
```

```
                              threshold=0.9)),
                 ('dfs', 'passthrough'),
                 ('pca',
                  Reduce_Dimensions_For_Supervised_Path(method='pca_liner',
                                                        random_state=123,
                                                        target='House price of '
                                                               'unit area',
        variance_retained_or_number_of_components=None))],
          verbose=False)
setup() succesfully completed…
```

[29]: `best_model = compare_models(sort='MAE')`

```
<pandas.io.formats.style.Styler at 0x23b92112bb0>

create_model_container: 19
master_model_container: 19
display_container: 2
ExtraTreesRegressor(bootstrap=False, ccp_alpha=0.0, criterion='mse',
                    max_depth=None, max_features='auto', max_leaf_nodes=None,
                    max_samples=None, min_impurity_decrease=0.0,
                    min_impurity_split=None, min_samples_leaf=1,
                    min_samples_split=2, min_weight_fraction_leaf=0.0,
                    n_estimators=100, n_jobs=-1, oob_score=False,
                    random_state=123, verbose=0, warm_start=False)
compare_models() succesfully completed…
```

[30]: `best_ = create_model(best_model)`

```
<pandas.io.formats.style.Styler at 0x23b9830f250>

create_model_container: 20
master_model_container: 20
display_container: 3
ExtraTreesRegressor(bootstrap=False, ccp_alpha=0.0, criterion='mse',
                    max_depth=None, max_features='auto', max_leaf_nodes=None,
                    max_samples=None, min_impurity_decrease=0.0,
                    min_impurity_split=None, min_samples_leaf=1,
                    min_samples_split=2, min_weight_fraction_leaf=0.0,
                    n_estimators=100, n_jobs=-1, oob_score=False,
                    random_state=123, verbose=0, warm_start=False)
create_model() succesfully completed…
```

[31]: `tuned_model = tune_model(best_)`

```
<pandas.io.formats.style.Styler at 0x23b97d75dc0>

create_model_container: 21
master_model_container: 21
display_container: 4
ExtraTreesRegressor(bootstrap=True, ccp_alpha=0.0, criterion='mse', max_depth=9,
```
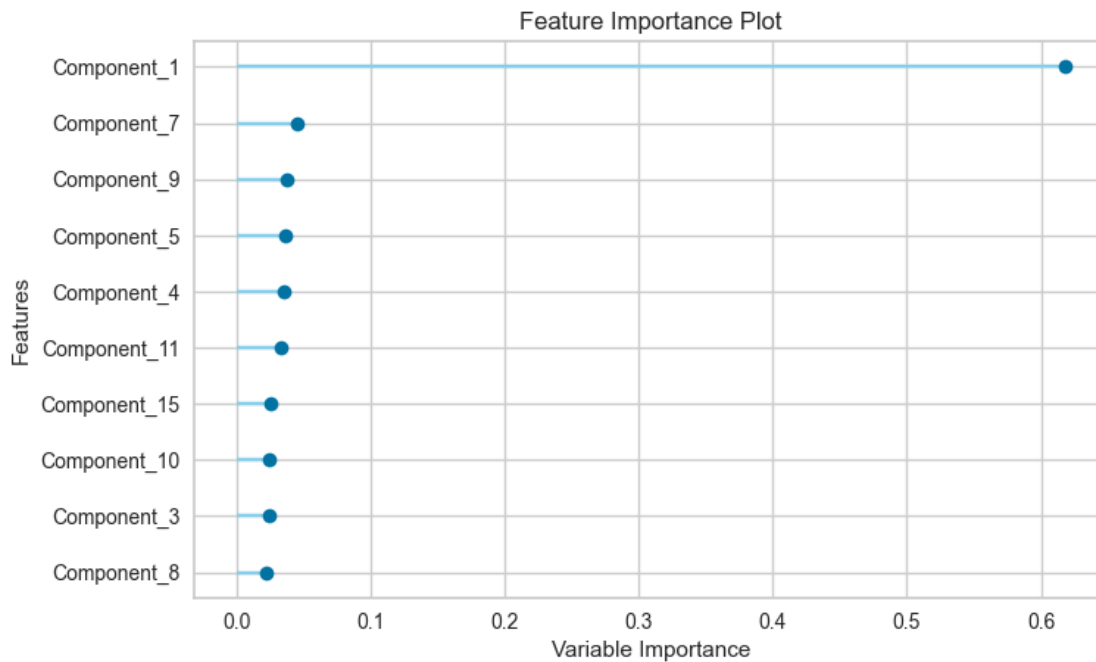
```
                    max_features=1.0, max_leaf_nodes=None, max_samples=None,
                    min_impurity_decrease=0.1, min_impurity_split=None,
                    min_samples_leaf=4, min_samples_split=7,
                    min_weight_fraction_leaf=0.0, n_estimators=100, n_jobs=-1,
                    oob_score=False, random_state=123, verbose=0,
                    warm_start=False)
tune_model() succesfully completed…
```

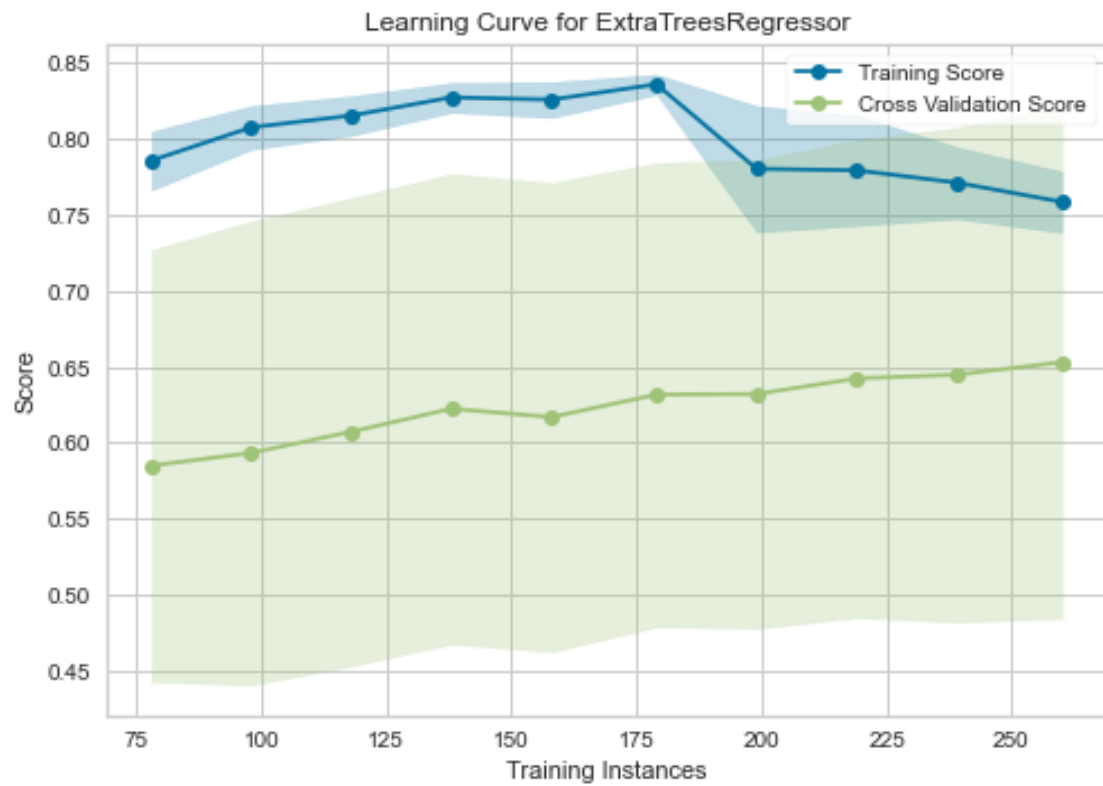[32]: `plot_model(estimator = tuned_model, plot = 'feature')`

Feature Importance Plot

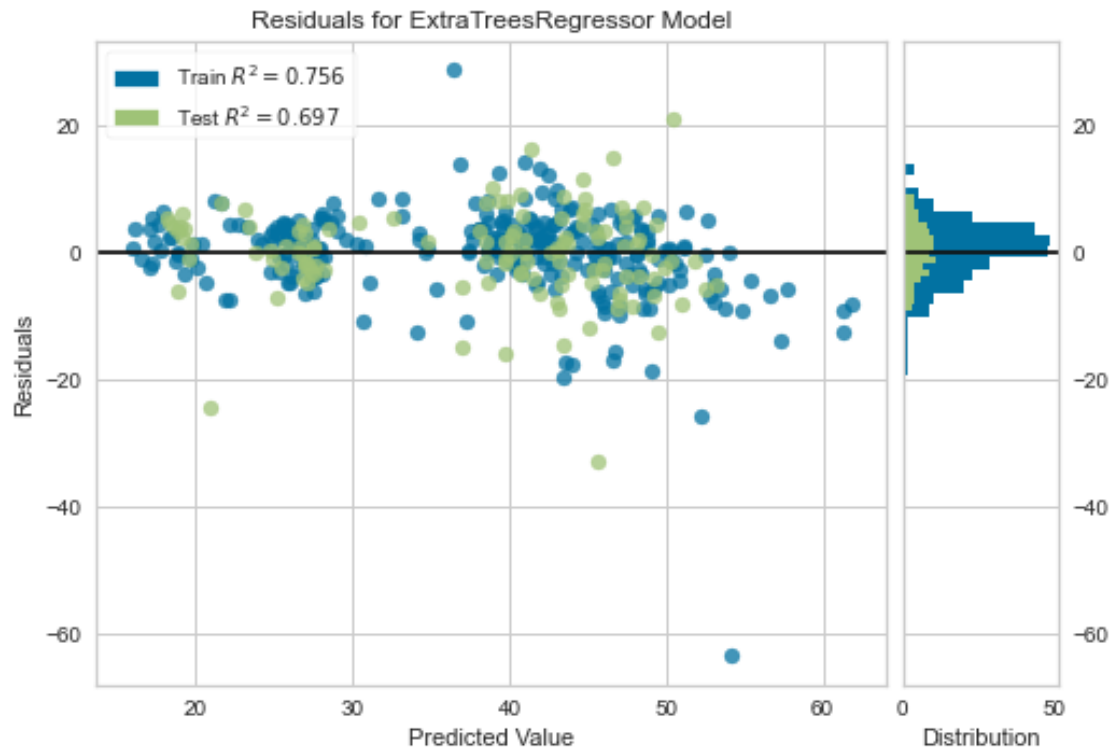```
Visual Rendered Successfully
plot_model() succesfully completed…
```

[33]: `plot_model(estimator = tuned_model, plot = 'learning')`

Visual Rendered Successfully
plot_model() succesfully completed…

[34]: `plot_model(tuned_model)`

Residuals for ExtraTreesRegressor Model

Train $R^2 = 0.756$
Test $R^2 = 0.697$

Residuals

Predicted Value

Distribution

```
Visual Rendered Successfully
plot_model() succesfully completed…
```

[ ]: