n6qnh4u08ZQPhYc

April 11, 2025

Install Libraries

[13]:
```
# importing modules and packages
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from matplotlib.legend_handler import HandlerPathCollection
from sklearn import preprocessing
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error
from sklearn.impute import SimpleImputer
from sklearn.neighbors import LocalOutlierFactor
from sklearn.model_selection import LearningCurveDisplay, ShuffleSplit
```

Read data from csv file

[14]:
```
# importing data
df = pd.read_csv('Car_Details_original.csv')
df.drop('Make', inplace = True, axis=1)

print(df.head())
print(df.columns)
```

```
    Type Origin DriveTrain  EngineSize  Cylinders  Horsepower  MPG_City  \
0    SUV   Asia        All         3.5        6.0         265        17
1  Sedan   Asia      Front         2.0        4.0         200        24
2  Sedan   Asia      Front         2.4        4.0         200        22
3  Sedan   Asia      Front         3.2        6.0         270        20
4  Sedan   Asia      Front         3.5        6.0         225        18

   MPG_Highway  Weight  Wheelbase  Length   MSRP
0           23    4451        106     189  36945
1           31    2778        101     172  23820
2           29    3230        105     183  26990
3           28    3575        108     186  33195
4           24    3880        115     197  43755
Index(['Type', 'Origin', 'DriveTrain', 'EngineSize', 'Cylinders', 'Horsepower',
```

```
 'MPG_City', 'MPG_Highway', 'Weight', 'Wheelbase', 'Length', 'MSRP'],
dtype='object')
```

Create X & Y variables then clean data.

[15]:
```
# creating feature variables
X = df.drop('MSRP', axis= 1)
x_array = np.array(X)
X_num = df[['EngineSize', 'Cylinders', 'Horsepower', 'MPG_City', 'MPG_Highway',␣
↪'Weight', 'Wheelbase', 'Length']]
xn_array = np.array(X_num)
print(X)
print(x_array)
print(xn_array)

y = df['MSRP']
y_array = np.array(y)
y_array = y_array.reshape(-1, 1)
print(y)
print(y_array)

# Encoding categorical features of X
enc = preprocessing.OrdinalEncoder()
enc.fit(x_array)
x_array = enc.transform(x_array)
print(x_array)

#Standardising X & Y
x_scaler = preprocessing.StandardScaler()
x_scaler.fit(xn_array)
print(x_scaler)
print(x_scaler.mean_)
print(x_scaler.scale_)
xn_array = x_scaler.transform(xn_array)
print(xn_array)

y_scaler = preprocessing.StandardScaler()
y_scaler.fit(y_array)
print(y_scaler)
print(y_scaler.mean_)
print(y_scaler.scale_)
y_array = y_scaler.transform(y_array)
print(y_array)
```

```
    Type  Origin DriveTrain  EngineSize  Cylinders  Horsepower  MPG_City  \
0    SUV    Asia        All         3.5        6.0         265        17
1  Sedan    Asia      Front         2.0        4.0         200        24
2  Sedan    Asia      Front         2.4        4.0         200        22
```

```
3    Sedan    Asia      Front         3.2         6.0         270         20
4    Sedan    Asia      Front         3.5         6.0         225         18
..     …      …         …         …         …         …       …
423  Sedan  Europe      Front         2.4         5.0         197         21
424  Sedan  Europe      Front         2.3         5.0         242         20
425  Sedan  Europe      Front         2.9         6.0         268         19
426  Wagon  Europe      Front         1.9         4.0         170         22
427  Wagon  Europe        All         2.5         5.0         208         20

     MPG_Highway  Weight  Wheelbase  Length
0             23    4451        106     189
1             31    2778        101     172
2             29    3230        105     183
3             28    3575        108     186
4             24    3880        115     197
..           …     …         …     …
423           28    3450        105     186
424           26    3450        105     186
425           26    3653        110     190
426           29    2822        101     180
427           27    3823        109     186

[428 rows x 11 columns]
[['SUV' 'Asia' 'All' … 4451 106 189]
 ['Sedan' 'Asia' 'Front' … 2778 101 172]
 ['Sedan' 'Asia' 'Front' … 3230 105 183]
 …
 ['Sedan' 'Europe' 'Front' … 3653 110 190]
 ['Wagon' 'Europe' 'Front' … 2822 101 180]
 ['Wagon' 'Europe' 'All' … 3823 109 186]]
[[3.500e+00 6.000e+00 2.650e+02 … 4.451e+03 1.060e+02 1.890e+02]
 [2.000e+00 4.000e+00 2.000e+02 … 2.778e+03 1.010e+02 1.720e+02]
 [2.400e+00 4.000e+00 2.000e+02 … 3.230e+03 1.050e+02 1.830e+02]
 …
 [2.900e+00 6.000e+00 2.680e+02 … 3.653e+03 1.100e+02 1.900e+02]
 [1.900e+00 4.000e+00 1.700e+02 … 2.822e+03 1.010e+02 1.800e+02]
 [2.500e+00 5.000e+00 2.080e+02 … 3.823e+03 1.090e+02 1.860e+02]]
0      36945
1      23820
2      26990
3      33195
4      43755
       …
423    40565
424    42565
425    45210
426    26135
427    35145
```

```
Name: MSRP, Length: 428, dtype: int64
[[ 36945]
 [ 23820]
 [ 26990]
 [ 33195]
 [ 43755]
 [ 46100]
 [ 89765]
 [ 25940]
 [ 35940]
 [ 31840]
 [ 33430]
 [ 34480]
 [ 36640]
 [ 39640]
 [ 42490]
 [ 44240]
 [ 42840]
 [ 49690]
 [ 69190]
 [ 48040]
 [ 84600]
 [ 35940]
 [ 37390]
 [ 40590]
 [ 40840]
 [ 49090]
 [ 37000]
 [ 52195]
 [ 28495]
 [ 30795]
 [ 37995]
 [ 30245]
 [ 35495]
 [ 36995]
 [ 37245]
 [ 39995]
 [ 44295]
 [ 44995]
 [ 54995]
 [ 69195]
 [ 73195]
 [ 48195]
 [ 56595]
 [ 33895]
 [ 41045]
 [ 32845]
 [ 37895]
```

```
[ 26545]
[ 22180]
[ 26470]
[ 24895]
[ 28345]
[ 32245]
[ 35545]
[ 40720]
[ 52795]
[ 46995]
[ 30835]
[ 45445]
[ 50595]
[ 47955]
[ 76200]
[ 52975]
[ 42735]
[ 41465]
[ 30295]
[ 20255]
[ 11690]
[ 12585]
[ 14610]
[ 14810]
[ 16385]
[ 21900]
[ 18995]
[ 20370]
[ 21825]
[ 25000]
[ 27995]
[ 23495]
[ 24225]
[ 26395]
[ 27020]
[ 44535]
[ 51535]
[ 36100]
[ 18760]
[ 20310]
[ 40340]
[ 41995]
[ 22225]
[ 17985]
[ 22000]
[ 19090]
[ 21840]
[ 29865]
```

```
[ 24130]
[ 26860]
[ 25955]
[ 25215]
[ 33295]
[ 30950]
[ 27490]
[ 38380]
[ 34495]
[ 31230]
[ 32235]
[ 13670]
[ 15040]
[ 22035]
[ 18820]
[ 20220]
[ 24885]
[ 21795]
[ 32660]
[ 81795]
[ 17630]
[ 20300]
[ 20215]
[ 41475]
[ 34560]
[ 29670]
[ 22515]
[ 13270]
[ 13730]
[ 15460]
[ 15580]
[ 19135]
[ 20320]
[ 22735]
[ 24345]
[ 27370]
[ 30315]
[ 26930]
[ 18345]
[ 29380]
[ 37530]
[ 22010]
[ 33540]
[ 14385]
[ 17475]
[ 22290]
[ 31890]
[ 35725]
```

[46265]
[25640]
[16530]
[25717]
[29322]
[25395]
[20140]
[19110]
[27560]
[19860]
[18690]
[13270]
[14170]
[15850]
[19860]
[22260]
[17750]
[19490]
[23760]
[26960]
[24950]
[27450]
[33260]
[49995]
[21589]
[10539]
[11839]
[11939]
[13839]
[15389]
[15389]
[19339]
[20339]
[24589]
[26189]
[18739]
[28495]
[29795]
[32445]
[31145]
[42845]
[52545]
[34895]
[36395]
[31849]
[20449]
[29995]
[33995]

```
[ 43895]
[ 49995]
[ 63120]
[ 68995]
[ 59995]
[ 74995]
[ 69995]
[ 74995]
[ 81995]
[ 86995]
[ 27905]
[ 20130]
[ 25520]
[ 19635]
[ 16040]
[ 10280]
[ 11155]
[ 12360]
[ 13580]
[ 14630]
[ 18435]
[ 26000]
[ 20615]
[ 11905]
[ 72250]
[ 39250]
[ 25995]
[ 45700]
[ 64800]
[ 39195]
[ 32350]
[ 31045]
[ 32415]
[ 41010]
[ 48450]
[ 55750]
[ 63200]
[ 32455]
[ 52775]
[ 42915]
[ 32495]
[ 36895]
[ 40095]
[ 43495]
[ 41815]
[ 44925]
[ 50470]
[ 16999]
```

```
[ 19999]
[ 21087]
[ 15500]
[ 17200]
[ 19270]
[ 28750]
[ 22388]
[ 25193]
[ 25700]
[ 27200]
[ 14840]
[ 22350]
[ 76870]
[ 46470]
[ 26060]
[ 28370]
[ 32280]
[ 33480]
[ 35920]
[ 37630]
[ 38830]
[ 52120]
[ 94820]
[128420]
[ 45707]
[ 52800]
[ 48170]
[ 57270]
[ 74320]
[ 86970]
[ 90520]
[121770]
[126670]
[ 40320]
[ 56170]
[ 33780]
[ 50670]
[ 60670]
[ 29995]
[ 21595]
[ 24695]
[ 29595]
[ 23895]
[ 30895]
[ 34495]
[ 33995]
[ 22595]
[ 30492]
```

```
[ 33112]
[ 18892]
[ 14622]
[ 16722]
[ 19312]
[ 17232]
[ 29282]
[ 25700]
[ 25092]
[ 26992]
[ 29562]
[ 17495]
[ 33840]
[ 27339]
[ 20939]
[ 12740]
[ 14740]
[ 19240]
[ 17640]
[ 23290]
[ 27490]
[ 29440]
[ 24780]
[ 32780]
[ 26910]
[ 34390]
[ 19479]
[ 26650]
[ 28739]
[ 18825]
[ 23675]
[ 28790]
[ 21595]
[ 15495]
[ 22450]
[ 22395]
[ 17735]
[ 24295]
[ 35995]
[ 23845]
[ 31370]
[ 33500]
[ 17045]
[ 56665]
[ 79165]
[ 84165]
[ 76765]
[192465]
```

```
[ 43365]
[ 52365]
[ 30860]
[ 33360]
[ 35105]
[ 39465]
[ 40670]
[ 43175]
[ 40845]
[ 20585]
[ 10995]
[ 14300]
[ 15825]
[ 14850]
[ 16350]
[ 21410]
[ 23560]
[ 12965]
[ 14165]
[ 19945]
[ 20445]
[ 25645]
[ 27145]
[ 29345]
[ 31545]
[ 25045]
[ 31545]
[ 24520]
[ 21445]
[ 23895]
[ 23699]
[ 17163]
[ 12884]
[ 14500]
[ 12269]
[ 15568]
[ 17262]
[ 16497]
[ 20510]
[ 35695]
[ 27710]
[ 27930]
[ 54765]
[ 20290]
[ 14085]
[ 15030]
[ 15295]
[ 10760]
```

```
 [ 11560]
 [ 11290]
 [ 19560]
 [ 22775]
 [ 19635]
 [ 21965]
 [ 26560]
 [ 25920]
 [ 26510]
 [ 30920]
 [ 23495]
 [ 28800]
 [ 22570]
 [ 25130]
 [ 12800]
 [ 16495]
 [ 25935]
 [ 16695]
 [ 35515]
 [ 18715]
 [ 19825]
 [ 21055]
 [ 21055]
 [ 23785]
 [ 23215]
 [ 23955]
 [ 33180]
 [ 39235]
 [ 65000]
 [ 75000]
 [ 19005]
 [ 24955]
 [ 40235]
 [ 41250]
 [ 25135]
 [ 31745]
 [ 34845]
 [ 37560]
 [ 37730]
 [ 37885]
 [ 40565]
 [ 42565]
 [ 45210]
 [ 26135]
 [ 35145]]
[[  1.   0.   0. … 300.  13.  37.]
 [  2.   0.   1. …  50.   8.  20.]
 [  2.   0.   1. … 105.  12.  31.]
```

```
 …
 [  2.   1.   1. … 201.  17.  38.]
 [  5.   1.   1. …  55.   8.  28.]
 [  5.   1.   0. … 228.  16.  34.]]
StandardScaler()
[3.19672897e+00 5.80751174e+00 2.15885514e+02 2.00607477e+01
 2.68434579e+01 3.57795327e+03 1.08154206e+02 1.86362150e+02]
[  1.10729887   1.5566124   71.75206191   5.23209465   5.73448978
 758.09603346   8.30209725  14.34120809]
[[ 0.27388362  0.12365844  0.68450278 …  1.15163078 -0.25947728
   0.18393502]
 [-1.08076419 -1.16118292 -0.22139453 … -1.05521363 -0.86173474
  -1.00146023]
 [-0.71952477 -1.16118292 -0.22139453 … -0.4589831  -0.37992877
  -0.23443977]
 …
 [-0.2679755   0.12365844  0.72631343 …  0.0989937   0.22232869
   0.25366416]
 [-1.17107405 -1.16118292 -0.63950098 … -0.99717349 -0.86173474
  -0.44362717]
 [-0.62921492 -0.51876224 -0.10989948 …  0.32323969  0.1018772
  -0.02525237]]
StandardScaler()
[32774.85514019]
[19409.00279492]
[[ 2.14856214e-01]
 [-4.61376364e-01]
 [-2.98050096e-01]
 [ 2.16469060e-02]
 [ 5.65724318e-01]
 [ 6.86544538e-01]
 [ 2.93627372e+00]
 [-3.52148702e-01]
 [ 1.63076120e-01]
 [-4.81660573e-02]
 [ 3.37546893e-02]
 [ 8.78532956e-02]
 [ 1.99141857e-01]
 [ 3.53709303e-01]
 [ 5.00548378e-01]
 [ 5.90712721e-01]
 [ 5.18581246e-01]
 [ 8.71510249e-01]
 [ 1.87619865e+00]
 [ 7.86498154e-01]
 [ 2.67016010e+00]
 [ 1.63076120e-01]
 [ 2.37783719e-01]
```

```
[ 4.02655662e-01]
[ 4.15536282e-01]
[ 8.40596760e-01]
[ 2.17689951e-01]
[ 1.00057407e+00]
[-2.20508760e-01]
[-1.02007051e-01]
[ 2.68954820e-01]
[-1.30344416e-01]
[ 1.40148615e-01]
[ 2.17432338e-01]
[ 2.30312959e-01]
[ 3.71999785e-01]
[ 5.93546458e-01]
[ 6.29612195e-01]
[ 1.14483702e+00]
[ 1.87645626e+00]
[ 2.08254619e+00]
[ 7.94484138e-01]
[ 1.22727299e+00]
[ 5.77126435e-02]
[ 4.26098391e-01]
[ 3.61403729e-03]
[ 2.63802572e-01]
[-3.20977600e-01]
[-5.45873235e-01]
[-3.24841786e-01]
[-4.05989696e-01]
[-2.28237132e-01]
[-2.72994520e-02]
[ 1.42724739e-01]
[ 4.09353584e-01]
[ 1.03148756e+00]
[ 7.32657160e-01]
[-9.99461518e-02]
[ 6.52797312e-01]
[ 9.18138095e-01]
[ 7.82118743e-01]
[ 2.23737125e+00]
[ 1.04076160e+00]
[ 5.13171386e-01]
[ 4.47737833e-01]
[-1.27768292e-01]
[-6.45054013e-01]
[-1.08634407e+00]
[-1.04023145e+00]
[-9.35898425e-01]
[-9.25593928e-01]
```

```
[-8.44446019e-01]
[-5.60299530e-01]
[-7.09972340e-01]
[-6.39128928e-01]
[-5.64163716e-01]
[-4.00579835e-01]
[-2.46270001e-01]
[-4.78121171e-01]
[-4.40509759e-01]
[-3.28705973e-01]
[-2.96504421e-01]
[ 6.05911854e-01]
[ 9.66569229e-01]
[ 1.71319717e-01]
[-7.22080124e-01]
[-6.42220276e-01]
[ 3.89775041e-01]
[ 4.75044749e-01]
[-5.43554723e-01]
[-7.62010047e-01]
[-5.55147282e-01]
[-7.05077705e-01]
[-5.63390879e-01]
[-1.49922960e-01]
[-4.45404395e-01]
[-3.04748018e-01]
[-3.51375865e-01]
[-3.89502501e-01]
[ 2.67991543e-02]
[-9.40210664e-02]
[-2.72288855e-01]
[ 2.88790976e-01]
[ 8.86261328e-02]
[-7.95947714e-02]
[-2.78146768e-02]
[-9.84329558e-01]
[-9.13743757e-01]
[-5.53343995e-01]
[-7.18988775e-01]
[-6.46857300e-01]
[-4.06504921e-01]
[-5.65709390e-01]
[-5.91762191e-03]
[ 2.52563954e+00]
[-7.80300529e-01]
[-6.42735501e-01]
[-6.47114912e-01]
[ 4.48253058e-01]
```

```
[ 9.19750942e-02]
[-1.59969844e-01]
[-5.28613203e-01]
[-1.00493855e+00]
[-9.81238209e-01]
[-8.92104315e-01]
[-8.85921617e-01]
[-7.02759193e-01]
[-6.41705052e-01]
[-5.17278257e-01]
[-4.34327061e-01]
[-2.78471552e-01]
[-1.26737843e-01]
[-3.01141445e-01]
[-7.43461954e-01]
[-1.74911363e-01]
[ 2.44996866e-01]
[-5.54632057e-01]
[ 3.94221624e-02]
[-9.47490983e-01]
[-7.88286513e-01]
[-5.40205762e-01]
[-4.55899332e-02]
[ 1.51998786e-01]
[ 6.95045748e-01]
[-3.67605447e-01]
[-8.36975259e-01]
[-3.63638215e-01]
[-1.77899667e-01]
[-3.80228455e-01]
[-6.50979098e-01]
[-7.04047255e-01]
[-2.68682281e-01]
[-6.65405393e-01]
[-7.25686698e-01]
[-1.00493855e+00]
[-9.58568317e-01]
[-8.72010547e-01]
[-6.65405393e-01]
[-5.41751436e-01]
[-7.74117831e-01]
[-6.84468712e-01]
[-4.64467713e-01]
[-2.99595770e-01]
[-4.03155959e-01]
[-2.74349754e-01]
[ 2.49958674e-02]
[ 8.87224606e-01]
```

```
[-5.76323022e-01]
[-1.14564645e+00]
[-1.07866722e+00]
[-1.07351497e+00]
[-9.75622258e-01]
[-8.95762411e-01]
[-8.95762411e-01]
[-6.92248607e-01]
[-6.40726124e-01]
[-4.21755575e-01]
[-3.39319604e-01]
[-7.23162096e-01]
[-2.20508760e-01]
[-1.53529533e-01]
[-1.69949556e-02]
[-8.39741824e-02]
[ 5.18838859e-01]
[ 1.01860694e+00]
[ 1.09235126e-01]
[ 1.86518849e-01]
[-4.77023549e-02]
[-6.35058651e-01]
[-1.43225037e-01]
[ 6.28648918e-02]
[ 5.72937465e-01]
[ 8.87224606e-01]
[ 1.56345718e+00]
[ 1.86615177e+00]
[ 1.40244943e+00]
[ 2.17528666e+00]
[ 1.91767425e+00]
[ 2.17528666e+00]
[ 2.53594403e+00]
[ 2.79355645e+00]
[-2.50907025e-01]
[-6.51494323e-01]
[-3.73788144e-01]
[-6.76997952e-01]
[-8.62221275e-01]
[-1.15899077e+00]
[-1.11390860e+00]
[-1.05182401e+00]
[-9.88966581e-01]
[-9.34867975e-01]
[-7.38824930e-01]
[-3.49057353e-01]
[-6.26505919e-01]
[-1.07526674e+00]
```

```
[ 2.03385745e+00]
[ 3.33615535e-01]
[-3.49314965e-01]
[ 6.65935545e-01]
[ 1.65001495e+00]
[ 3.30781799e-01]
[-2.18895914e-02]
[-8.91264306e-02]
[-1.85406300e-02]
[ 4.24295104e-01]
[ 8.07622371e-01]
[ 1.18373649e+00]
[ 1.56757898e+00]
[-1.64797308e-02]
[ 1.03045711e+00]
[ 5.22445433e-01]
[-1.44188315e-02]
[ 2.12280090e-01]
[ 3.77152033e-01]
[ 5.52328472e-01]
[ 4.65770702e-01]
[ 6.26005622e-01]
[ 9.11697785e-01]
[-8.12811215e-01]
[-6.58243768e-01]
[-6.02187308e-01]
[-8.90043416e-01]
[-8.02455196e-01]
[-6.95803658e-01]
[-2.07370527e-01]
[-5.35156559e-01]
[-3.90635996e-01]
[-3.64514098e-01]
[-2.87230374e-01]
[-9.24048254e-01]
[-5.37114413e-01]
[ 2.27189131e+00]
[ 7.05607857e-01]
[-3.45966004e-01]
[-2.26949070e-01]
[-2.54961651e-02]
[ 3.63308135e-02]
[ 1.62045670e-01]
[ 2.50149114e-01]
[ 3.11976093e-01]
[ 9.96709881e-01]
[ 3.19671987e+00]
[ 4.92787527e+00]
```

```
[ 6.66296203e-01]
[ 1.03174517e+00]
[ 7.93196076e-01]
[ 1.26205066e+00]
[ 2.14050898e+00]
[ 2.79226838e+00]
[ 2.97517320e+00]
[ 4.58525076e+00]
[ 4.83771092e+00]
[ 3.88744591e-01]
[ 1.20537593e+00]
[ 5.17875581e-02]
[ 9.22002282e-01]
[ 1.43722710e+00]
[-1.43225037e-01]
[-5.76013887e-01]
[-4.16294192e-01]
[-1.63834030e-01]
[-4.57512178e-01]
[-9.68548029e-02]
[ 8.86261328e-02]
[ 6.28648918e-02]
[-5.24491405e-01]
[-1.17618363e-01]
[ 1.73705400e-02]
[-7.15279156e-01]
[-9.35280155e-01]
[-8.27082942e-01]
[-6.93639714e-01]
[-8.00806476e-01]
[-1.79960567e-01]
[-3.64514098e-01]
[-3.95839767e-01]
[-2.97947051e-01]
[-1.65534272e-01]
[-7.87256064e-01]
[ 5.48789070e-02]
[-2.80068749e-01]
[-6.09812635e-01]
[-1.03224547e+00]
[-9.29200502e-01]
[-6.97349332e-01]
[-7.79785304e-01]
[-4.88683280e-01]
[-2.72288855e-01]
[-1.71820014e-01]
[-4.11914781e-01]
[ 2.65075948e-04]
```

```
[-3.02171894e-01]
[ 8.32162722e-02]
[-6.85035459e-01]
[-3.15567740e-01]
[-2.07937274e-01]
[-7.18731162e-01]
[-4.68847124e-01]
[-2.05309628e-01]
[-5.76013887e-01]
[-8.90301028e-01]
[-5.31962165e-01]
[-5.34795901e-01]
[-7.74890668e-01]
[-4.36903185e-01]
[ 1.65909856e-01]
[-4.60088302e-01]
[-7.23816239e-02]
[ 3.73612631e-02]
[-8.10441181e-01]
[ 1.23087956e+00]
[ 2.39013541e+00]
[ 2.64774782e+00]
[ 2.26648145e+00]
[ 8.22763264e+00]
[ 5.45630549e-01]
[ 1.00933289e+00]
[-9.86580898e-02]
[ 3.01481156e-02]
[ 1.20054847e-01]
[ 3.44692869e-01]
[ 4.06777460e-01]
[ 5.35841278e-01]
[ 4.15793894e-01]
[-6.28051594e-01]
[-1.12215220e+00]
[-9.51870394e-01]
[-8.73298609e-01]
[-9.23533029e-01]
[-8.46249306e-01]
[-5.85545546e-01]
[-4.74772209e-01]
[-1.02065291e+00]
[-9.58825929e-01]
[-6.61025982e-01]
[-6.35264741e-01]
[-3.67347834e-01]
[-2.90064111e-01]
[-1.76714650e-01]
```

```
[-6.33651895e-02]
[-3.98261323e-01]
[-6.33651895e-02]
[-4.25310627e-01]
[-5.83742259e-01]
[-4.57512178e-01]
[-4.67610584e-01]
[-8.04361528e-01]
[-1.02482623e+00]
[-9.41565898e-01]
[-1.05651256e+00]
[-8.86539887e-01]
[-7.99260802e-01]
[-8.38675501e-01]
[-6.31915780e-01]
[ 1.50453111e-01]
[-2.60953909e-01]
[-2.49618962e-01]
[ 1.13298685e+00]
[-6.43250726e-01]
[-9.62947728e-01]
[-9.14258982e-01]
[-9.00605524e-01]
[-1.13425998e+00]
[-1.09304200e+00]
[-1.10695307e+00]
[-6.80862138e-01]
[-5.15217358e-01]
[-6.76997952e-01]
[-5.56950568e-01]
[-3.20204763e-01]
[-3.53179152e-01]
[-3.22780887e-01]
[-9.55667408e-02]
[-4.78121171e-01]
[-2.04794403e-01]
[-5.25779467e-01]
[-3.93881912e-01]
[-1.02915412e+00]
[-8.38778546e-01]
[-3.52406314e-01]
[-8.28474049e-01]
[ 1.41179065e-01]
[-7.24398635e-01]
[-6.67208680e-01]
[-6.03836027e-01]
[-6.03836027e-01]
[-4.63179651e-01]
```

```
[-4.92547466e-01]
[-4.54420829e-01]
[ 2.08740688e-02]
[ 3.32842698e-01]
[ 1.66031945e+00]
[ 2.17554427e+00]
[-7.09457116e-01]
[-4.02898347e-01]
[ 3.84365180e-01]
[ 4.36660500e-01]
[-3.93624300e-01]
[-5.30606931e-02]
[ 1.06659002e-01]
[ 2.46542541e-01]
[ 2.55301363e-01]
[ 2.63287347e-01]
[ 4.01367599e-01]
[ 5.04412564e-01]
[ 6.40689529e-01]
[-3.42101818e-01]
[ 1.22115746e-01]]
```

Detecting Missing Values and Outliers. Plot current heatmap.

[16]:
```
# Handle Missing text values in X variables.
x_imp = SimpleImputer(strategy="most_frequent")
x_imp.fit(x_array)
x_array = x_imp.transform(x_array)
print(x_array)

# Handle NaN values in X and Y variables.
xn_imp = SimpleImputer(missing_values=np.nan, strategy='mean')
xn_imp.fit(xn_array)
xn_array = xn_imp.transform(xn_array)
print(xn_array)

y_imp = SimpleImputer(missing_values=np.nan, strategy='mean')
y_imp.fit(y_array)
y_array = y_imp.transform(y_array)
print(y_array)

# Detect Outliers
ground_truth = np.ones(len(xn_array), dtype=int)
clf = LocalOutlierFactor(n_neighbors=20, contamination=0.1)
y_pred = clf.fit_predict(xn_array)
n_errors = (y_pred != ground_truth).sum()
X_scores = clf.negative_outlier_factor_
```

```python
# Plot current heatmap
def update_legend_marker_size(handle, orig):
    "Customize size of the legend marker"
    handle.update_from(orig)
    handle.set_sizes([20])

plt.scatter(xn_array[:, 0], xn_array[:, 1], color="k", s=3.0, label="Data␣
↪points")
# plot circles with radius proportional to the outlier scores
radius = (X_scores.max() - X_scores) / (X_scores.max() - X_scores.min())
scatter = plt.scatter(
    xn_array[:, 0],
    xn_array[:, 1],
    s=1000 * radius,
    edgecolors="r",
    facecolors="none",
    label="Outlier scores",
)
plt.axis("tight")
plt.xlim((-5, 10))
plt.ylim((-5, 10))
plt.xlabel("prediction errors: %d" % (n_errors))
plt.legend(
    handler_map={scatter:␣
↪HandlerPathCollection(update_func=update_legend_marker_size)}
)
plt.title("Local Outlier Factor (LOF)")
plt.show()
```

```
[[  1.    0.    0. … 300.   13.   37.]
 [  2.    0.    1. …  50.    8.   20.]
 [  2.    0.    1. … 105.   12.   31.]
 …
 [  2.    1.    1. … 201.   17.   38.]
 [  5.    1.    1. …  55.    8.   28.]
 [  5.    1.    0. … 228.   16.   34.]]
[[ 0.27388362  0.12365844  0.68450278 …  1.15163078 -0.25947728
   0.18393502]
 [-1.08076419 -1.16118292 -0.22139453 … -1.05521363 -0.86173474
  -1.00146023]
 [-0.71952477 -1.16118292 -0.22139453 … -0.4589831  -0.37992877
  -0.23443977]
 …
 [-0.2679755   0.12365844  0.72631343 …  0.0989937   0.22232869
   0.25366416]
 [-1.17107405 -1.16118292 -0.63950098 … -0.99717349 -0.86173474
  -0.44362717]]
```

```
 [-0.62921492 -0.51876224 -0.10989948 …  0.32323969  0.1018772
  -0.02525237]]
[[ 2.14856214e-01]
 [-4.61376364e-01]
 [-2.98050096e-01]
 [ 2.16469060e-02]
 [ 5.65724318e-01]
 [ 6.86544538e-01]
 [ 2.93627372e+00]
 [-3.52148702e-01]
 [ 1.63076120e-01]
 [-4.81660573e-02]
 [ 3.37546893e-02]
 [ 8.78532956e-02]
 [ 1.99141857e-01]
 [ 3.53709303e-01]
 [ 5.00548378e-01]
 [ 5.90712721e-01]
 [ 5.18581246e-01]
 [ 8.71510249e-01]
 [ 1.87619865e+00]
 [ 7.86498154e-01]
 [ 2.67016010e+00]
 [ 1.63076120e-01]
 [ 2.37783719e-01]
 [ 4.02655662e-01]
 [ 4.15536282e-01]
 [ 8.40596760e-01]
 [ 2.17689951e-01]
 [ 1.00057407e+00]
 [-2.20508760e-01]
 [-1.02007051e-01]
 [ 2.68954820e-01]
 [-1.30344416e-01]
 [ 1.40148615e-01]
 [ 2.17432338e-01]
 [ 2.30312959e-01]
 [ 3.71999785e-01]
 [ 5.93546458e-01]
 [ 6.29612195e-01]
 [ 1.14483702e+00]
 [ 1.87645626e+00]
 [ 2.08254619e+00]
 [ 7.94484138e-01]
 [ 1.22727299e+00]
 [ 5.77126435e-02]
 [ 4.26098391e-01]
 [ 3.61403729e-03]
```

```
[ 2.63802572e-01]
[-3.20977600e-01]
[-5.45873235e-01]
[-3.24841786e-01]
[-4.05989696e-01]
[-2.28237132e-01]
[-2.72994520e-02]
[ 1.42724739e-01]
[ 4.09353584e-01]
[ 1.03148756e+00]
[ 7.32657160e-01]
[-9.99461518e-02]
[ 6.52797312e-01]
[ 9.18138095e-01]
[ 7.82118743e-01]
[ 2.23737125e+00]
[ 1.04076160e+00]
[ 5.13171386e-01]
[ 4.47737833e-01]
[-1.27768292e-01]
[-6.45054013e-01]
[-1.08634407e+00]
[-1.04023145e+00]
[-9.35898425e-01]
[-9.25593928e-01]
[-8.44446019e-01]
[-5.60299530e-01]
[-7.09972340e-01]
[-6.39128928e-01]
[-5.64163716e-01]
[-4.00579835e-01]
[-2.46270001e-01]
[-4.78121171e-01]
[-4.40509759e-01]
[-3.28705973e-01]
[-2.96504421e-01]
[ 6.05911854e-01]
[ 9.66569229e-01]
[ 1.71319717e-01]
[-7.22080124e-01]
[-6.42220276e-01]
[ 3.89775041e-01]
[ 4.75044749e-01]
[-5.43554723e-01]
[-7.62010047e-01]
[-5.55147282e-01]
[-7.05077705e-01]
[-5.63390879e-01]
```

```
[-1.49922960e-01]
[-4.45404395e-01]
[-3.04748018e-01]
[-3.51375865e-01]
[-3.89502501e-01]
[ 2.67991543e-02]
[-9.40210664e-02]
[-2.72288855e-01]
[ 2.88790976e-01]
[ 8.86261328e-02]
[-7.95947714e-02]
[-2.78146768e-02]
[-9.84329558e-01]
[-9.13743757e-01]
[-5.53343995e-01]
[-7.18988775e-01]
[-6.46857300e-01]
[-4.06504921e-01]
[-5.65709390e-01]
[-5.91762191e-03]
[ 2.52563954e+00]
[-7.80300529e-01]
[-6.42735501e-01]
[-6.47114912e-01]
[ 4.48253058e-01]
[ 9.19750942e-02]
[-1.59969844e-01]
[-5.28613203e-01]
[-1.00493855e+00]
[-9.81238209e-01]
[-8.92104315e-01]
[-8.85921617e-01]
[-7.02759193e-01]
[-6.41705052e-01]
[-5.17278257e-01]
[-4.34327061e-01]
[-2.78471552e-01]
[-1.26737843e-01]
[-3.01141445e-01]
[-7.43461954e-01]
[-1.74911363e-01]
[ 2.44996866e-01]
[-5.54632057e-01]
[ 3.94221624e-02]
[-9.47490983e-01]
[-7.88286513e-01]
[-5.40205762e-01]
[-4.55899332e-02]
```

```
[ 1.51998786e-01]
[ 6.95045748e-01]
[-3.67605447e-01]
[-8.36975259e-01]
[-3.63638215e-01]
[-1.77899667e-01]
[-3.80228455e-01]
[-6.50979098e-01]
[-7.04047255e-01]
[-2.68682281e-01]
[-6.65405393e-01]
[-7.25686698e-01]
[-1.00493855e+00]
[-9.58568317e-01]
[-8.72010547e-01]
[-6.65405393e-01]
[-5.41751436e-01]
[-7.74117831e-01]
[-6.84468712e-01]
[-4.64467713e-01]
[-2.99595770e-01]
[-4.03155959e-01]
[-2.74349754e-01]
[ 2.49958674e-02]
[ 8.87224606e-01]
[-5.76323022e-01]
[-1.14564645e+00]
[-1.07866722e+00]
[-1.07351497e+00]
[-9.75622258e-01]
[-8.95762411e-01]
[-8.95762411e-01]
[-6.92248607e-01]
[-6.40726124e-01]
[-4.21755575e-01]
[-3.39319604e-01]
[-7.23162096e-01]
[-2.20508760e-01]
[-1.53529533e-01]
[-1.69949556e-02]
[-8.39741824e-02]
[ 5.18838859e-01]
[ 1.01860694e+00]
[ 1.09235126e-01]
[ 1.86518849e-01]
[-4.77023549e-02]
[-6.35058651e-01]
[-1.43225037e-01]
```

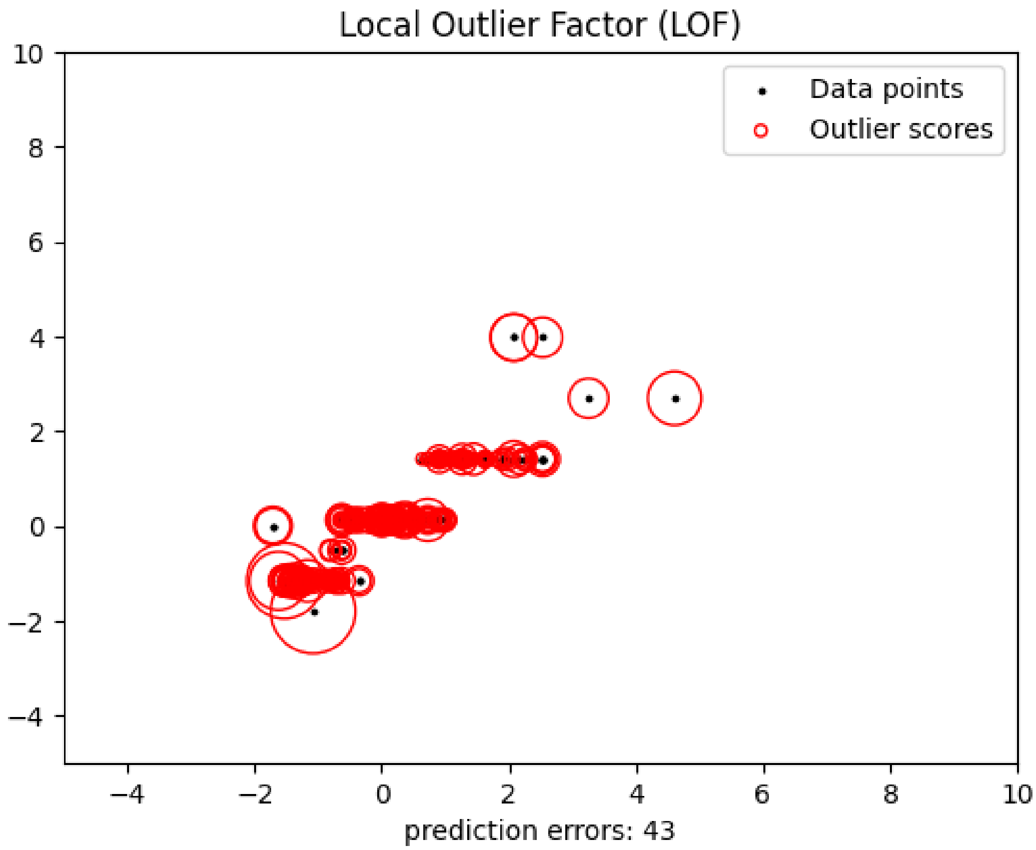
```
[ 6.28648918e-02]
[ 5.72937465e-01]
[ 8.87224606e-01]
[ 1.56345718e+00]
[ 1.86615177e+00]
[ 1.40244943e+00]
[ 2.17528666e+00]
[ 1.91767425e+00]
[ 2.17528666e+00]
[ 2.53594403e+00]
[ 2.79355645e+00]
[-2.50907025e-01]
[-6.51494323e-01]
[-3.73788144e-01]
[-6.76997952e-01]
[-8.62221275e-01]
[-1.15899077e+00]
[-1.11390860e+00]
[-1.05182401e+00]
[-9.88966581e-01]
[-9.34867975e-01]
[-7.38824930e-01]
[-3.49057353e-01]
[-6.26505919e-01]
[-1.07526674e+00]
[ 2.03385745e+00]
[ 3.33615535e-01]
[-3.49314965e-01]
[ 6.65935545e-01]
[ 1.65001495e+00]
[ 3.30781799e-01]
[-2.18895914e-02]
[-8.91264306e-02]
[-1.85406300e-02]
[ 4.24295104e-01]
[ 8.07622371e-01]
[ 1.18373649e+00]
[ 1.56757898e+00]
[-1.64797308e-02]
[ 1.03045711e+00]
[ 5.22445433e-01]
[-1.44188315e-02]
[ 2.12280090e-01]
[ 3.77152033e-01]
[ 5.52328472e-01]
[ 4.65770702e-01]
[ 6.26005622e-01]
[ 9.11697785e-01]
```

```
[-8.12811215e-01]
[-6.58243768e-01]
[-6.02187308e-01]
[-8.90043416e-01]
[-8.02455196e-01]
[-6.95803658e-01]
[-2.07370527e-01]
[-5.35156559e-01]
[-3.90635996e-01]
[-3.64514098e-01]
[-2.87230374e-01]
[-9.24048254e-01]
[-5.37114413e-01]
[ 2.27189131e+00]
[ 7.05607857e-01]
[-3.45966004e-01]
[-2.26949070e-01]
[-2.54961651e-02]
[ 3.63308135e-02]
[ 1.62045670e-01]
[ 2.50149114e-01]
[ 3.11976093e-01]
[ 9.96709881e-01]
[ 3.19671987e+00]
[ 4.92787527e+00]
[ 6.66296203e-01]
[ 1.03174517e+00]
[ 7.93196076e-01]
[ 1.26205066e+00]
[ 2.14050898e+00]
[ 2.79226838e+00]
[ 2.97517320e+00]
[ 4.58525076e+00]
[ 4.83771092e+00]
[ 3.88744591e-01]
[ 1.20537593e+00]
[ 5.17875581e-02]
[ 9.22002282e-01]
[ 1.43722710e+00]
[-1.43225037e-01]
[-5.76013887e-01]
[-4.16294192e-01]
[-1.63834030e-01]
[-4.57512178e-01]
[-9.68548029e-02]
[ 8.86261328e-02]
[ 6.28648918e-02]
[-5.24491405e-01]
```

```
[-1.17618363e-01]
[ 1.73705400e-02]
[-7.15279156e-01]
[-9.35280155e-01]
[-8.27082942e-01]
[-6.93639714e-01]
[-8.00806476e-01]
[-1.79960567e-01]
[-3.64514098e-01]
[-3.95839767e-01]
[-2.97947051e-01]
[-1.65534272e-01]
[-7.87256064e-01]
[ 5.48789070e-02]
[-2.80068749e-01]
[-6.09812635e-01]
[-1.03224547e+00]
[-9.29200502e-01]
[-6.97349332e-01]
[-7.79785304e-01]
[-4.88683280e-01]
[-2.72288855e-01]
[-1.71820014e-01]
[-4.11914781e-01]
[ 2.65075948e-04]
[-3.02171894e-01]
[ 8.32162722e-02]
[-6.85035459e-01]
[-3.15567740e-01]
[-2.07937274e-01]
[-7.18731162e-01]
[-4.68847124e-01]
[-2.05309628e-01]
[-5.76013887e-01]
[-8.90301028e-01]
[-5.31962165e-01]
[-5.34795901e-01]
[-7.74890668e-01]
[-4.36903185e-01]
[ 1.65909856e-01]
[-4.60088302e-01]
[-7.23816239e-02]
[ 3.73612631e-02]
[-8.10441181e-01]
[ 1.23087956e+00]
[ 2.39013541e+00]
[ 2.64774782e+00]
[ 2.26648145e+00]
```

```
[ 8.22763264e+00]
[ 5.45630549e-01]
[ 1.00933289e+00]
[-9.86580898e-02]
[ 3.01481156e-02]
[ 1.20054847e-01]
[ 3.44692869e-01]
[ 4.06777460e-01]
[ 5.35841278e-01]
[ 4.15793894e-01]
[-6.28051594e-01]
[-1.12215220e+00]
[-9.51870394e-01]
[-8.73298609e-01]
[-9.23533029e-01]
[-8.46249306e-01]
[-5.85545546e-01]
[-4.74772209e-01]
[-1.02065291e+00]
[-9.58825929e-01]
[-6.61025982e-01]
[-6.35264741e-01]
[-3.67347834e-01]
[-2.90064111e-01]
[-1.76714650e-01]
[-6.33651895e-02]
[-3.98261323e-01]
[-6.33651895e-02]
[-4.25310627e-01]
[-5.83742259e-01]
[-4.57512178e-01]
[-4.67610584e-01]
[-8.04361528e-01]
[-1.02482623e+00]
[-9.41565898e-01]
[-1.05651256e+00]
[-8.86539887e-01]
[-7.99260802e-01]
[-8.38675501e-01]
[-6.31915780e-01]
[ 1.50453111e-01]
[-2.60953909e-01]
[-2.49618962e-01]
[ 1.13298685e+00]
[-6.43250726e-01]
[-9.62947728e-01]
[-9.14258982e-01]
[-9.00605524e-01]
```

```
[-1.13425998e+00]
[-1.09304200e+00]
[-1.10695307e+00]
[-6.80862138e-01]
[-5.15217358e-01]
[-6.76997952e-01]
[-5.56950568e-01]
[-3.20204763e-01]
[-3.53179152e-01]
[-3.22780887e-01]
[-9.55667408e-02]
[-4.78121171e-01]
[-2.04794403e-01]
[-5.25779467e-01]
[-3.93881912e-01]
[-1.02915412e+00]
[-8.38778546e-01]
[-3.52406314e-01]
[-8.28474049e-01]
[ 1.41179065e-01]
[-7.24398635e-01]
[-6.67208680e-01]
[-6.03836027e-01]
[-6.03836027e-01]
[-4.63179651e-01]
[-4.92547466e-01]
[-4.54420829e-01]
[ 2.08740688e-02]
[ 3.32842698e-01]
[ 1.66031945e+00]
[ 2.17554427e+00]
[-7.09457116e-01]
[-4.02898347e-01]
[ 3.84365180e-01]
[ 4.36660500e-01]
[-3.93624300e-01]
[-5.30606931e-02]
[ 1.06659002e-01]
[ 2.46542541e-01]
[ 2.55301363e-01]
[ 2.63287347e-01]
[ 4.01367599e-01]
[ 5.04412564e-01]
[ 6.40689529e-01]
[-3.42101818e-01]
[ 1.22115746e-01]]
```



Create a Linear Regression model

[17]:
```
# creating train and test sets
X_train, X_test, y_train, y_test = train_test_split(xn_array, y_array,␣
 ↪test_size=0.3, random_state=101)

# creating a regression model
model = LinearRegression()

# fitting the model
model.fit(X_train,y_train)

# making predictions
predictions = model.predict(X_test)
print(predictions)

# model evaluation
print('Test mean_squared_error : ', mean_squared_error(y_test, predictions))
print( 'Test mean_absolute_error : ', mean_absolute_error(y_test, predictions))
```

```
[[-5.84400016e-01]
```

```
[ 1.32254861e-01]
[ 3.27772812e-01]
[-6.05447767e-01]
[-5.30924351e-01]
[ 5.57157163e-03]
[ 2.20833707e-01]
[-4.98926267e-01]
[-2.69332324e-01]
[-8.15650135e-01]
[-6.17320953e-01]
[-8.72374775e-01]
[-8.97804204e-01]
[-1.14428234e+00]
[ 9.62466467e-01]
[ 1.40432254e+00]
[-1.13387324e+00]
[ 5.60176917e-01]
[ 3.30903599e-02]
[-2.86362311e-01]
[-3.12094012e-01]
[-2.00835818e-01]
[-3.09483596e-01]
[ 1.29332117e+00]
[ 5.59249270e-01]
[-1.55699473e-01]
[ 1.69603935e+00]
[ 4.10522313e-02]
[-9.82505428e-01]
[-5.44439518e-01]
[ 1.92166687e-01]
[ 1.56061208e+00]
[ 8.86217028e-02]
[ 6.93148539e-01]
[ 7.85415977e-01]
[-1.13335131e+00]
[-8.58099306e-01]
[-7.97406408e-01]
[-3.18175553e-01]
[ 4.33215370e-01]
[ 2.81923712e-01]
[-3.67050990e-01]
[ 9.13548203e-01]
[-9.69742968e-01]
[-8.24356639e-01]
[ 1.01793539e+00]
[ 5.36588239e-02]
[ 2.21609503e-01]
[-2.74060359e-01]
```

```
[-3.57446165e-01]
[ 1.07985887e-01]
[-2.59053403e-02]
[ 4.89562256e-01]
[-1.40088302e-01]
[ 9.69804743e-01]
[-1.30009677e-01]
[ 2.50129809e-01]
[-8.68567761e-02]
[-1.15943391e+00]
[ 4.30799753e-01]
[-8.15650135e-01]
[ 1.66302008e-01]
[ 2.25422151e-03]
[ 1.17279632e+00]
[ 2.03082822e-01]
[-7.32194136e-01]
[-1.69615207e-01]
[-9.84639905e-01]
[ 1.14781384e+00]
[-4.23731733e-01]
[-7.75996348e-01]
[-2.37564667e-01]
[-4.70231143e-01]
[-7.49726080e-01]
[ 2.61058181e-01]
[-8.89869795e-02]
[ 2.38970713e-01]
[-7.97406408e-01]
[-2.31582664e-02]
[-1.08570366e+00]
[-2.96408062e-01]
[ 1.03526418e+00]
[-9.81238532e-01]
[ 4.12422435e+00]
[-7.93715674e-01]
[-4.89255196e-01]
[ 1.17517253e+00]
[ 5.73532440e-01]
[-4.63089421e-01]
[ 1.31977847e+00]
[ 3.15123297e-01]
[ 1.50202445e-01]
[ 1.73681068e-01]
[ 2.40342911e-02]
[ 1.57870500e+00]
[ 2.54767614e-02]
[ 1.97043878e-01]
```

```
 [ 1.20794462e+00]
 [-9.98205123e-01]
 [ 3.02108274e-01]
 [ 4.39202613e-01]
 [ 8.61620751e-01]
 [-7.78786400e-01]
 [-5.58699717e-01]
 [-6.99867311e-01]
 [-6.15101377e-01]
 [ 8.18825178e-01]
 [-7.85762239e-01]
 [-4.52841904e-01]
 [-2.85261961e-01]
 [-9.26211908e-01]
 [-8.34749799e-01]
 [ 2.96909440e-01]
 [ 1.58550861e-01]
 [-1.16829904e+00]
 [ 1.81711086e-01]
 [-1.15422845e-01]
 [ 8.18825178e-01]
 [ 4.98486926e-03]
 [-6.54949509e-01]
 [ 8.31910779e-01]
 [-7.58058673e-01]
 [-6.95045002e-01]
 [ 1.34654272e+00]
 [-1.28547730e-01]
 [ 3.22264892e+00]
 [ 7.11765018e-03]
 [ 2.59840532e-01]
 [-1.60923069e-01]]
Test mean_squared_error :  0.2190712772454837
Test mean_absolute_error :  0.3404897404251003
```

Plot learning curve for model.

[18]:
```python
# Create the regression model
regressor = RandomForestRegressor(n_estimators=100, random_state=42)
fig, ax = plt.subplots(nrows=1, ncols=2, figsize=(10, 6), sharey=True)

common_params = {
    "X": xn_array,
    "y": y_array,
    "train_sizes": np.linspace(0.1, 1.0, 5),
    "cv": ShuffleSplit(n_splits=50, test_size=0.2, random_state=0),
    "score_type": "both",
    "n_jobs": 4,
```

```
    "line_kw": {"marker": "o"},
    "std_display_style": "fill_between",
    "score_name": "R2 Score",
}

# Create a list of estimators to compare
estimators = [model, regressor]

for ax_idx, estimator in enumerate(estimators):
    LearningCurveDisplay.from_estimator(estimator, **common_params,␣
 ↪ax=ax[ax_idx])
    handles, label = ax[ax_idx].get_legend_handles_labels()
    ax[ax_idx].legend(handles[:2], ["Training Score", "Test Score"])
    ax[ax_idx].set_title(f"Learning Curve for {estimator.__class__.__name__}")
```