IBM_Python_Cars

December 17, 2024

[ ]:
```
IBM PYTHON MACHINE LEARNING COURSE
MODULE: LOADING AND EXPLORING DATA
```

[1]:
```python
# import pandas library
import pandas as pd
import numpy as np
```

[48]:
```python
import pandas as pd
import os

# Define the file path
file_path = r"IBM_Python_Car_Dataset.csv"

# Check if the file exists
if os.path.exists(file_path):
    # Read the CSV file with explicit delimiter (comma) and check the first few␣
 ↪rows
    data = pd.read_csv(file_path, encoding='utf-8', header=None, delimiter=',')

    # Inspect the first few rows to check if columns are split correctly
    print("First few rows of the dataset:\n", data.head())

    # Check the shape of the dataset (number of rows and columns)
    print("Shape of the dataset:", data.shape)

    # If data only has one column, try using `sep` instead of `delimiter`
    if data.shape[1] == 1:
        data = pd.read_csv(file_path, encoding='utf-8', header=None,␣
 ↪sep=r'\s*,\s*', engine='python')  # Regular expression for comma

    # Define the column names (adjust the list size based on the dataset's␣
 ↪columns)
    column_names = ['symboling', 'normalized-losses', 'make', 'fuel-type',␣
 ↪'aspiration', 'num-of-doors',
                    'body-style', 'drive-wheels', 'engine-location',␣
 ↪'wheel-base', 'length', 'width',
```

```
                     'height', 'curb-weight', 'engine-type', 'num-of-cylinders',␣
↪'engine-size', 'fuel-system',
                     'bore', 'stroke', 'compression-ratio', 'horsepower',␣
↪'peak-rpm', 'city-mpg', 'highway-mpg',
                     'price']

    # Assign the column names to the dataset, if the number of columns match
    if data.shape[1] == len(column_names):
        data.columns = column_names
        print("Dataset with column names:\n", data.head())
    else:
        print(f"Warning: Column mismatch. Expected {len(column_names)} columns,␣
↪but got {data.shape[1]}")
else:
    print(f"File not found: {file_path}")
```

```
First few rows of the dataset:
                                                    0
0  3,?,alfa-romero,gas,std,two,convertible,rwd,fr…
1  3,?,alfa-romero,gas,std,two,convertible,rwd,fr…
2  1,?,alfa-romero,gas,std,two,hatchback,rwd,fron…
3  2,164,audi,gas,std,four,sedan,fwd,front,99.80,…
4  2,164,audi,gas,std,four,sedan,4wd,front,99.40,…
Shape of the dataset: (205, 1)
Dataset with column names:
   symboling normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3                 ?  alfa-romero       gas        std          two
1        "3                 ?  alfa-romero       gas        std          two
2        "1                 ?  alfa-romero       gas        std          two
3        "2               164         audi       gas        std         four
4        "2               164         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
0  convertible          rwd           front        88.6  …          130
1  convertible          rwd           front        88.6  …          130
2    hatchback          rwd           front        94.5  …          152
3        sedan          fwd           front        99.8  …          109
4        sedan          4wd           front        99.4  …          136

   fuel-system  bore   stroke compression-ratio horsepower  peak-rpm city-mpg  \
0         mpfi  3.47     2.68               9.0        111      5000       21
1         mpfi  3.47     2.68               9.0        111      5000       21
2         mpfi  2.68     3.47               9.0        154      5000       19
3         mpfi  3.19     3.40              10.0        102      5500       24
4         mpfi  3.19     3.40               8.0        115      5500       18

  highway-mpg   price
```

```
0           27  13495"
1           27  16500"
2           26  16500"
3           30  13950"
4           22  17450"

[5 rows x 26 columns]
```

EXPLORE DATA SET

[50]:
```python
# Display the first 5 rows of the dataset
print(data.head())
```

```
  symboling normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3                 ?  alfa-romero       gas        std          two
1        "3                 ?  alfa-romero       gas        std          two
2        "1                 ?  alfa-romero       gas        std          two
3        "2               164         audi       gas        std         four
4        "2               164         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
0  convertible          rwd           front        88.6  …          130
1  convertible          rwd           front        88.6  …          130
2    hatchback          rwd           front        94.5  …          152
3        sedan          fwd           front        99.8  …          109
4        sedan          4wd           front        99.4  …          136

   fuel-system  bore  stroke compression-ratio horsepower  peak-rpm city-mpg  \
0         mpfi  3.47    2.68               9.0        111      5000       21
1         mpfi  3.47    2.68               9.0        111      5000       21
2         mpfi  2.68    3.47               9.0        154      5000       19
3         mpfi  3.19    3.40              10.0        102      5500       24
4         mpfi  3.19    3.40               8.0        115      5500       18

  highway-mpg   price
0           27  13495"
1           27  16500"
2           26  16500"
3           30  13950"
4           22  17450"

[5 rows x 26 columns]
```

[52]:
```python
# Display the last 5 rows of the dataset
print(data.tail())
```

```
    symboling normalized-losses   make fuel-type aspiration num-of-doors  \
200        "-1                95  volvo       gas        std         four
201        "-1                95  volvo       gas      turbo         four
```

```
202       "-1                 95  volvo       gas       std         four
203       "-1                 95  volvo    diesel     turbo         four
204       "-1                 95  volvo       gas     turbo         four

    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
200      sedan          rwd           front       109.1  …          141
201      sedan          rwd           front       109.1  …          141
202      sedan          rwd           front       109.1  …          173
203      sedan          rwd           front       109.1  …          145
204      sedan          rwd           front       109.1  …          141

     fuel-system  bore  stroke compression-ratio horsepower  peak-rpm  \
200         mpfi  3.78    3.15               9.5        114      5400
201         mpfi  3.78    3.15               8.7        160      5300
202         mpfi  3.58    2.87               8.8        134      5500
203          idi  3.01    3.40              23.0        106      4800
204         mpfi  3.78    3.15               9.5        114      5400

    city-mpg highway-mpg   price
200       23          28  16845"
201       19          25  19045"
202       18          23  21485"
203       26          27  22470"
204       19          25  22625"

[5 rows x 26 columns]
```

WHEN YOU TAKE A BREAK YOU HAVE TO RESTART THE SESSION

[20]:
```python
import pandas as pd

# Example: Loading a dataset from a CSV file
data = pd.read_csv('IBM_Python_Car_Dataset.csv')  # Replace 'your_dataset.csv'
 ↪with your file path
```

[22]:
```python
# Display dataset information
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 204 entries, 0 to 203
Data columns (total 1 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   3,?,alfa-romero,gas,std,two,convertible,rwd,front,88.60,168.80,64.10,48.80,
2548,dohc,four,130,mpfi,3.47,2.68,9.00,111,5000,21,27,13495  204 non-null
object
```

```
dtypes: object(1)
memory usage: 1.7+ KB
```

[24]:
```
# Display the first 5 rows of the dataset
print(data.head())
```

```
  3,?,alfa-romero,gas,std,two,convertible,rwd,front,88.60,168.80,64.10,48.80,254
8,dohc,four,130,mpfi,3.47,2.68,9.00,111,5000,21,27,13495
0  3,?,alfa-romero,gas,std,two,convertible,rwd,fr…
1  1,?,alfa-romero,gas,std,two,hatchback,rwd,fron…
2  2,164,audi,gas,std,four,sedan,fwd,front,99.80,…
3  2,164,audi,gas,std,four,sedan,4wd,front,99.40,…
4  2,?,audi,gas,std,two,sedan,fwd,front,99.80,177…
```

THE WORK WAS NOT SAVED BECAUSE YOU ENDED THE SESSION W/O CREATING A NEW CSV

[26]:
```
import pandas as pd
import os

# Define the file path
file_path = r"IBM_Python_Car_Dataset.csv"

# Check if the file exists
if os.path.exists(file_path):
    # Read the CSV file with explicit delimiter (comma) and check the first few␣
 ↪rows
    data = pd.read_csv(file_path, encoding='utf-8', header=None, delimiter=',')

    # Inspect the first few rows to check if columns are split correctly
    print("First few rows of the dataset:\n", data.head())

    # Check the shape of the dataset (number of rows and columns)
    print("Shape of the dataset:", data.shape)

    # If data only has one column, try using `sep` instead of `delimiter`
    if data.shape[1] == 1:
        data = pd.read_csv(file_path, encoding='utf-8', header=None,␣
 ↪sep=r'\s*,\s*', engine='python')  # Regular expression for comma

    # Define the column names (adjust the list size based on the dataset's␣
 ↪columns)
    column_names = ['symboling', 'normalized-losses', 'make', 'fuel-type',␣
 ↪'aspiration', 'num-of-doors',
                    'body-style', 'drive-wheels', 'engine-location',␣
 ↪'wheel-base', 'length', 'width',
                    'height', 'curb-weight', 'engine-type', 'num-of-cylinders',␣
 ↪'engine-size', 'fuel-system',
```

```
                    'bore', 'stroke', 'compression-ratio', 'horsepower',␣
↪'peak-rpm', 'city-mpg', 'highway-mpg',
                    'price']

    # Assign the column names to the dataset, if the number of columns match
    if data.shape[1] == len(column_names):
        data.columns = column_names
        print("Dataset with column names:\n", data.head())
    else:
        print(f"Warning: Column mismatch. Expected {len(column_names)} columns,␣
↪but got {data.shape[1]}")
else:
    print(f"File not found: {file_path}")
```

```
First few rows of the dataset:
                                                    0
0  3,?,alfa-romero,gas,std,two,convertible,rwd,fr…
1  3,?,alfa-romero,gas,std,two,convertible,rwd,fr…
2  1,?,alfa-romero,gas,std,two,hatchback,rwd,fron…
3  2,164,audi,gas,std,four,sedan,fwd,front,99.80,…
4  2,164,audi,gas,std,four,sedan,4wd,front,99.40,…
Shape of the dataset: (205, 1)
Dataset with column names:
   symboling normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3                 ?  alfa-romero       gas        std          two
1        "3                 ?  alfa-romero       gas        std          two
2        "1                 ?  alfa-romero       gas        std          two
3        "2               164         audi       gas        std         four
4        "2               164         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
0  convertible          rwd           front        88.6  …          130
1  convertible          rwd           front        88.6  …          130
2    hatchback          rwd           front        94.5  …          152
3        sedan          fwd           front        99.8  …          109
4        sedan          4wd           front        99.4  …          136

   fuel-system  bore  stroke compression-ratio horsepower  peak-rpm city-mpg  \
0         mpfi  3.47    2.68               9.0        111      5000       21
1         mpfi  3.47    2.68               9.0        111      5000       21
2         mpfi  2.68    3.47               9.0        154      5000       19
3         mpfi  3.19    3.40              10.0        102      5500       24
4         mpfi  3.19    3.40               8.0        115      5500       18

  highway-mpg   price
0          27  13495"
1          27  16500"
```

```
2           26  16500"
3           30  13950"
4           22  17450"

[5 rows x 26 columns]
```

[29]:
```
# Display the first 5 rows of the dataset
print(data.head())
```

```
  symboling normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3                 ?  alfa-romero       gas        std          two
1        "3                 ?  alfa-romero       gas        std          two
2        "1                 ?  alfa-romero       gas        std          two
3        "2               164         audi       gas        std         four
4        "2               164         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
0  convertible          rwd           front        88.6  …          130
1  convertible          rwd           front        88.6  …          130
2    hatchback          rwd           front        94.5  …          152
3        sedan          fwd           front        99.8  …          109
4        sedan          4wd           front        99.4  …          136

   fuel-system  bore  stroke compression-ratio horsepower  peak-rpm city-mpg  \
0         mpfi  3.47    2.68               9.0        111      5000       21
1         mpfi  3.47    2.68               9.0        111      5000       21
2         mpfi  2.68    3.47               9.0        154      5000       19
3         mpfi  3.19    3.40              10.0        102      5500       24
4         mpfi  3.19    3.40               8.0        115      5500       18

  highway-mpg   price
0          27  13495"
1          27  16500"
2          26  16500"
3          30  13950"
4          22  17450"

[5 rows x 26 columns]
```

SAVE UPDATED FILE TO ANOTHER CSV

[31]:
```
data.to_csv(r'C:\Users\dj1975\Documents\LinearRegres.csv', index=False)
```

VERIFY NEW CSV EXISTS

[33]:
```
import os

filepath = r'C:\Users\dj1975\Documents\LinearRegres.csv'
print(os.path.exists(filepath))  # Prints True if the file exists
```

```
True
```

[35]:
```
# Display the first 5 rows of the dataset
print(data.head())
```

```
  symboling normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3                 ?  alfa-romero       gas        std          two
1        "3                 ?  alfa-romero       gas        std          two
2        "1                 ?  alfa-romero       gas        std          two
3        "2               164         audi       gas        std         four
4        "2               164         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
0  convertible          rwd           front        88.6  …          130
1  convertible          rwd           front        88.6  …          130
2    hatchback          rwd           front        94.5  …          152
3        sedan          fwd           front        99.8  …          109
4        sedan          4wd           front        99.4  …          136

   fuel-system  bore  stroke compression-ratio horsepower  peak-rpm city-mpg  \
0         mpfi  3.47    2.68               9.0        111      5000       21
1         mpfi  3.47    2.68               9.0        111      5000       21
2         mpfi  2.68    3.47               9.0        154      5000       19
3         mpfi  3.19    3.40              10.0        102      5500       24
4         mpfi  3.19    3.40               8.0        115      5500       18

  highway-mpg   price
0          27  13495"
1          27  16500"
2          26  16500"
3          30  13950"
4          22  17450"

[5 rows x 26 columns]
```

[39]:
```
# Display dataset information
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 205 entries, 0 to 204
Data columns (total 26 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   symboling          205 non-null    object
 1   normalized-losses  205 non-null    object
 2   make               205 non-null    object
 3   fuel-type          205 non-null    object
 4   aspiration         205 non-null    object
```

```
 5   num-of-doors       205 non-null    object
 6   body-style         205 non-null    object
 7   drive-wheels       205 non-null    object
 8   engine-location    205 non-null    object
 9   wheel-base         205 non-null    float64
 10  length             205 non-null    float64
 11  width              205 non-null    float64
 12  height             205 non-null    float64
 13  curb-weight        205 non-null    int64
 14  engine-type        205 non-null    object
 15  num-of-cylinders   205 non-null    object
 16  engine-size        205 non-null    int64
 17  fuel-system        205 non-null    object
 18  bore               205 non-null    object
 19  stroke             205 non-null    object
 20  compression-ratio  205 non-null    float64
 21  horsepower         205 non-null    object
 22  peak-rpm           205 non-null    object
 23  city-mpg           205 non-null    int64
 24  highway-mpg        205 non-null    int64
 25  price              205 non-null    object
dtypes: float64(5), int64(4), object(17)
memory usage: 41.8+ KB
```

[49]:
```
data.describe()
```

[49]:
```
       wheel-base      length       width      height  curb-weight  \
count  205.000000  205.000000  205.000000  205.000000   205.000000
mean    98.756585  174.049268   65.907805   53.724878  2555.565854
std      6.021776   12.337289    2.145204    2.443522   520.680204
min     86.600000  141.100000   60.300000   47.800000  1488.000000
25%     94.500000  166.300000   64.100000   52.000000  2145.000000
50%     97.000000  173.200000   65.500000   54.100000  2414.000000
75%    102.400000  183.100000   66.900000   55.500000  2935.000000
max    120.900000  208.100000   72.300000   59.800000  4066.000000

       engine-size  compression-ratio    city-mpg  highway-mpg
count   205.000000         205.000000  205.000000   205.000000
mean    126.907317          10.142537   25.219512    30.751220
std      41.642693           3.972040    6.542142     6.886443
min      61.000000           7.000000   13.000000    16.000000
25%      97.000000           8.600000   19.000000    25.000000
50%     120.000000           9.000000   24.000000    30.000000
75%     141.000000           9.400000   30.000000    34.000000
max     326.000000          23.000000   49.000000    54.000000
```

[57]:
```
data.describe(include='all')
```

[57]:

```
       symboling normalized-losses    make fuel-type aspiration num-of-doors  \
count        205               205     205       205        205          205
unique         6                52      22         2          2            3
top           "0                 ?  toyota       gas        std         four
freq          67                41      32       185        168          114
mean         NaN               NaN     NaN       NaN        NaN          NaN
std          NaN               NaN     NaN       NaN        NaN          NaN
min          NaN               NaN     NaN       NaN        NaN          NaN
25%          NaN               NaN     NaN       NaN        NaN          NaN
50%          NaN               NaN     NaN       NaN        NaN          NaN
75%          NaN               NaN     NaN       NaN        NaN          NaN
max          NaN               NaN     NaN       NaN        NaN          NaN

       body-style drive-wheels engine-location  wheel-base  …  engine-size  \
count         205          205             205  205.000000  …   205.000000
unique          5            3               2         NaN  …          NaN
top         sedan          fwd           front         NaN  …          NaN
freq           96          120             202         NaN  …          NaN
mean          NaN          NaN             NaN   98.756585  …   126.907317
std           NaN          NaN             NaN    6.021776  …    41.642693
min           NaN          NaN             NaN   86.600000  …    61.000000
25%           NaN          NaN             NaN   94.500000  …    97.000000
50%           NaN          NaN             NaN   97.000000  …   120.000000
75%           NaN          NaN             NaN  102.400000  …   141.000000
max           NaN          NaN             NaN  120.900000  …   326.000000

        fuel-system  bore  stroke compression-ratio horsepower  peak-rpm  \
count           205   205     205        205.000000        205       205
unique            8    39      37               NaN         60        24
top            mpfi  3.62    3.40               NaN         68      5500
freq             94    23      20               NaN         19        37
mean            NaN   NaN     NaN         10.142537        NaN       NaN
std             NaN   NaN     NaN          3.972040        NaN       NaN
min             NaN   NaN     NaN          7.000000        NaN       NaN
25%             NaN   NaN     NaN          8.600000        NaN       NaN
50%             NaN   NaN     NaN          9.000000        NaN       NaN
75%             NaN   NaN     NaN          9.400000        NaN       NaN
max             NaN   NaN     NaN         23.000000        NaN       NaN

          city-mpg highway-mpg price
count   205.000000  205.000000   205
unique         NaN         NaN   187
top            NaN         NaN    ?"
freq           NaN         NaN     4
mean     25.219512   30.751220   NaN
std       6.542142    6.886443   NaN
min      13.000000   16.000000   NaN
```

```
25%        19.000000   25.000000   NaN
50%        24.000000   30.000000   NaN
75%        30.000000   34.000000   NaN
max        49.000000   54.000000   NaN

[11 rows x 26 columns]
```

MODULE 2: DATA PRE-PROCESSING

DEALING WITH MISSING VALUES

[59]:
```
#Drop Missing Data in "Price" Column
data = data.dropna(subset=["price"], axis=0)
```

[63]:
```
#Check row count before and after #dropna command
print("Before dropna:", data.shape)
data = data.dropna(subset=["price"], axis=0)
print("After dropna:", data.shape)
```

```
Before dropna: (205, 26)
After dropna: (205, 26)
```

[80]:
```
# Convert any non-numeric values to NaN
df["normalized-losses"] = pd.to_numeric(df["normalized-losses"],␣
↪errors='coerce')
```

[82]:
```
# Calculate the mean of the "normalized-losses" column (ignoring NaN values)
mean_value = df["normalized-losses"].mean()
```

[84]:
```
# Replace missing values (NaN) in the "normalized-losses" column with the␣
↪calculated mean
df["normalized-losses"] = df["normalized-losses"].fillna(mean_value)
```

[88]:
```
# Check if there are any remaining NaN values in the 'normalized-losses' column
print(df["normalized-losses"].isna().sum())  # This will print the count of NaN␣
↪values

# Optionally, display the first few rows to inspect the 'normalized-losses'␣
↪column
print(df.head())
```

```
0
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3               122.0  alfa-romero       gas        std          two
1        "3               122.0  alfa-romero       gas        std          two
2        "1               122.0  alfa-romero       gas        std          two
3        "2               164.0         audi       gas        std         four
4        "2               164.0         audi       gas        std         four
```

```
    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
0  convertible          rwd           front        88.6  …          130
1  convertible          rwd           front        88.6  …          130
2    hatchback          rwd           front        94.5  …          152
3        sedan          fwd           front        99.8  …          109
4        sedan          4wd           front        99.4  …          136

   fuel-system  bore  stroke compression-ratio horsepower  peak-rpm city-mpg  \
0         mpfi  3.47    2.68               9.0        111      5000       21
1         mpfi  3.47    2.68               9.0        111      5000       21
2         mpfi  2.68    3.47               9.0        154      5000       19
3         mpfi  3.19    3.40              10.0        102      5500       24
4         mpfi  3.19    3.40               8.0        115      5500       18

  highway-mpg   price
0          27  13495"
1          27  16500"
2          26  16500"
3          30  13950"
4          22  17450"

[5 rows x 26 columns]
```

[90]:
```python
# Convert city-mpg to City-L/100km
df["city-mpg"] = 235 / df["city-mpg"]
```

[108]:
```python
# Rename the column
df.rename(columns = {"City-L/100km": "city-L/100km"}, inplace=True)
```

[110]:
```python
# Check the first few rows of the dataframe to see if the transformation and␣
↪renaming worked
print(df.head())
```

```
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3              122.0  alfa-romero       gas        std          two
1        "3              122.0  alfa-romero       gas        std          two
2        "1              122.0  alfa-romero       gas        std          two
3        "2              164.0         audi       gas        std         four
4        "2              164.0         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
0  convertible          rwd           front        88.6  …          130
1  convertible          rwd           front        88.6  …          130
2    hatchback          rwd           front        94.5  …          152
3        sedan          fwd           front        99.8  …          109
4        sedan          4wd           front        99.4  …          136
```

```
   fuel-system  bore  stroke compression-ratio horsepower  peak-rpm  \
0         mpfi  3.47    2.68               9.0        111      5000
1         mpfi  3.47    2.68               9.0        111      5000
2         mpfi  2.68    3.47               9.0        154      5000
3         mpfi  3.19    3.40              10.0        102      5500
4         mpfi  3.19    3.40               8.0        115      5500

  city-L/100km highway-mpg   price
0    11.190476          27  13495"
1    11.190476          27  16500"
2    12.368421          26  16500"
3     9.791667          30  13950"
4    13.055556          22  17450"

[5 rows x 26 columns]
```

[112]:
```python
# Verify the column names
print(df.columns)
```

```
Index(['symboling', 'normalized-losses', 'make', 'fuel-type', 'aspiration',
       'num-of-doors', 'body-style', 'drive-wheels', 'engine-location',
       'wheel-base', 'length', 'width', 'height', 'curb-weight', 'engine-type',
       'num-of-cylinders', 'engine-size', 'fuel-system', 'bore', 'stroke',
       'compression-ratio', 'horsepower', 'peak-rpm', 'city-L/100km',
       'highway-mpg', 'price'],
      dtype='object')
```

[114]:
```python
# Save the updated DataFrame to the original CSV file
df.to_csv(r'C:\Users\dj1975\Documents\LinearRegres.csv', index=False)
```

[118]:
```python
import datetime

# Convert the timestamp to a human-readable format
timestamp = 1732638970.839595
readable_time = datetime.datetime.fromtimestamp(timestamp)

# Print the formatted time
print(readable_time)
```

```
2024-11-26 16:36:10.839595
```

[123]:
```python
# Remove non-numeric characters (like the double quotes) from the 'price' column
df["price"] = df["price"].replace({r'[^\d]': ''}, regex=True)
```

[127]:
```python
# Remove rows with empty strings or non-numeric values in 'price' column
df = df[df['price'].apply(lambda x: str(x).isdigit())]
```

[129]:
```python
# Remove non-numeric characters (like the double quotes) from the 'price' column
df["price"] = df["price"].replace({r'[^\d]': ''}, regex=True)
```

[131]:
```python
# Now convert the 'price' column to integers
df["price"] = df["price"].astype(int)
```

[137]:
```python
# Display general information about the DataFrame
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 201 entries, 0 to 204
Data columns (total 26 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   symboling          201 non-null    object
 1   normalized-losses  201 non-null    float64
 2   make               201 non-null    object
 3   fuel-type          201 non-null    object
 4   aspiration         201 non-null    object
 5   num-of-doors       201 non-null    object
 6   body-style         201 non-null    object
 7   drive-wheels       201 non-null    object
 8   engine-location    201 non-null    object
 9   wheel-base         201 non-null    float64
 10  length             201 non-null    float64
 11  width              201 non-null    float64
 12  height             201 non-null    float64
 13  curb-weight        201 non-null    int64
 14  engine-type        201 non-null    object
 15  num-of-cylinders   201 non-null    object
 16  engine-size        201 non-null    int64
 17  fuel-system        201 non-null    object
 18  bore               201 non-null    object
 19  stroke             201 non-null    object
 20  compression-ratio  201 non-null    float64
 21  horsepower         201 non-null    object
 22  peak-rpm           201 non-null    object
 23  city-L/100km       201 non-null    float64
 24  highway-mpg        201 non-null    int64
 25  price              201 non-null    int64
dtypes: float64(7), int64(4), object(15)
memory usage: 42.4+ KB
```

RE=STARTING AFTER A BREAK

[146]:
```python
import pandas as pd

# Load the CSV file into a DataFrame
```

```
df = pd.read_csv(r'C:\Users\dj1975\Documents\LinearRegres.csv')

# Verify the DataFrame is loaded
print(df.head())  # Show the first 5 rows to confirm the data
```

```
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3              122.0  alfa-romero       gas        std          two
1        "3              122.0  alfa-romero       gas        std          two
2        "1              122.0  alfa-romero       gas        std          two
3        "2              164.0         audi       gas        std         four
4        "2              164.0         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
0  convertible          rwd           front        88.6  …          130
1  convertible          rwd           front        88.6  …          130
2    hatchback          rwd           front        94.5  …          152
3        sedan          fwd           front        99.8  …          109
4        sedan          4wd           front        99.4  …          136

   fuel-system  bore  stroke compression-ratio horsepower  peak-rpm  \
0         mpfi  3.47    2.68               9.0        111      5000
1         mpfi  3.47    2.68               9.0        111      5000
2         mpfi  2.68    3.47               9.0        154      5000
3         mpfi  3.19    3.40              10.0        102      5500
4         mpfi  3.19    3.40               8.0        115      5500

  city-L/100km highway-mpg  price
0    11.190476          27  13495
1    11.190476          27  16500
2    12.368421          26  16500
3     9.791667          30  13950
4    13.055556          22  17450

[5 rows x 26 columns]
```

[148]:
```
# Calculate the mean of the "normalized-losses" column
mean = df["normalized-losses"].mean()
```

[150]:
```
# Replace NaN values in the "normalized-losses" column with the calculated mean
df["normalized-losses"] = df["normalized-losses"].replace(np.nan, mean)
```

[152]:
```
# Verify the changes
print(df["normalized-losses"].isna().sum())  # Should print 0 if all NaNs are
 ↪replaced
```

```
0
```

[154]:
```python
# Verify the DataFrame is loaded
print(df.head())  # Show the first 5 rows to confirm the data
```

```
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3              122.0  alfa-romero       gas        std          two
1        "3              122.0  alfa-romero       gas        std          two
2        "1              122.0  alfa-romero       gas        std          two
3        "2              164.0         audi       gas        std         four
4        "2              164.0         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
0  convertible          rwd           front        88.6  …          130
1  convertible          rwd           front        88.6  …          130
2    hatchback          rwd           front        94.5  …          152
3        sedan          fwd           front        99.8  …          109
4        sedan          4wd           front        99.4  …          136

   fuel-system  bore  stroke compression-ratio horsepower  peak-rpm  \
0         mpfi  3.47    2.68               9.0        111      5000
1         mpfi  3.47    2.68               9.0        111      5000
2         mpfi  2.68    3.47               9.0        154      5000
3         mpfi  3.19    3.40              10.0        102      5500
4         mpfi  3.19    3.40               8.0        115      5500

  city-L/100km highway-mpg  price
0    11.190476          27  13495
1    11.190476          27  16500
2    12.368421          26  16500
3     9.791667          30  13950
4    13.055556          22  17450

[5 rows x 26 columns]
```

[156]:
```python
# Check for NaN values in the dataset
print(df.isnull().sum())
```

```
symboling            0
normalized-losses    0
make                 0
fuel-type            0
aspiration           0
num-of-doors         0
body-style           0
drive-wheels         0
engine-location      0
wheel-base           0
length               0
width                0
```

```
height                0
curb-weight           0
engine-type           0
num-of-cylinders      0
engine-size           0
fuel-system           0
bore                  0
stroke                0
compression-ratio     0
horsepower            0
peak-rpm              0
city-L/100km          0
highway-mpg           0
price                 0
dtype: int64
```

[158]:
```
print(df.head(50))  # Show 50 rows to confirm the data
```

```
   symboling  normalized-losses         make fuel-type aspiration  \
0         "3              122.0  alfa-romero       gas        std
1         "3              122.0  alfa-romero       gas        std
2         "1              122.0  alfa-romero       gas        std
3         "2              164.0         audi       gas        std
4         "2              164.0         audi       gas        std
5         "2              122.0         audi       gas        std
6         "1              158.0         audi       gas        std
7         "1              122.0         audi       gas        std
8         "1              158.0         audi       gas      turbo
9         "2              192.0          bmw       gas        std
10        "0              192.0          bmw       gas        std
11        "0              188.0          bmw       gas        std
12        "0              188.0          bmw       gas        std
13        "1              122.0          bmw       gas        std
14        "0              122.0          bmw       gas        std
15        "0              122.0          bmw       gas        std
16        "0              122.0          bmw       gas        std
17        "2              121.0    chevrolet       gas        std
18        "1               98.0    chevrolet       gas        std
19        "0               81.0    chevrolet       gas        std
20        "1              118.0        dodge       gas        std
21        "1              118.0        dodge       gas        std
22        "1              118.0        dodge       gas      turbo
23        "1              148.0        dodge       gas        std
24        "1              148.0        dodge       gas        std
25        "1              148.0        dodge       gas        std
26        "1              148.0        dodge       gas      turbo
27       "-1              110.0        dodge       gas        std
28        "3              145.0        dodge       gas      turbo
```

```
29         "2              137.0        honda       gas        std
30         "2              137.0        honda       gas        std
31         "1              101.0        honda       gas        std
32         "1              101.0        honda       gas        std
33         "1              101.0        honda       gas        std
34         "0              110.0        honda       gas        std
35         "0               78.0        honda       gas        std
36         "0              106.0        honda       gas        std
37         "0              106.0        honda       gas        std
38         "0               85.0        honda       gas        std
39         "0               85.0        honda       gas        std
40         "0               85.0        honda       gas        std
41         "1              107.0        honda       gas        std
42         "0              122.0        isuzu       gas        std
43         "2              122.0        isuzu       gas        std
44         "0              145.0       jaguar       gas        std
45         "0              122.0       jaguar       gas        std
46         "0              122.0       jaguar       gas        std
47         "1              104.0        mazda       gas        std
48         "1              104.0        mazda       gas        std
49         "1              104.0        mazda       gas        std

   num-of-doors   body-style drive-wheels engine-location  wheel-base  …  \
0           two  convertible          rwd           front        88.6  …
1           two  convertible          rwd           front        88.6  …
2           two    hatchback          rwd           front        94.5  …
3          four        sedan          fwd           front        99.8  …
4          four        sedan          4wd           front        99.4  …
5           two        sedan          fwd           front        99.8  …
6          four        sedan          fwd           front       105.8  …
7          four        wagon          fwd           front       105.8  …
8          four        sedan          fwd           front       105.8  …
9           two        sedan          rwd           front       101.2  …
10         four        sedan          rwd           front       101.2  …
11          two        sedan          rwd           front       101.2  …
12         four        sedan          rwd           front       101.2  …
13         four        sedan          rwd           front       103.5  …
14         four        sedan          rwd           front       103.5  …
15          two        sedan          rwd           front       103.5  …
16         four        sedan          rwd           front       110.0  …
17          two    hatchback          fwd           front        88.4  …
18          two    hatchback          fwd           front        94.5  …
19         four        sedan          fwd           front        94.5  …
20          two    hatchback          fwd           front        93.7  …
21          two    hatchback          fwd           front        93.7  …
22          two    hatchback          fwd           front        93.7  …
23         four    hatchback          fwd           front        93.7  …
24         four        sedan          fwd           front        93.7  …
```

```
25          four        sedan          fwd           front        93.7  …
26             ?        sedan          fwd           front        93.7  …
27          four        wagon          fwd           front       103.3  …
28           two    hatchback          fwd           front        95.9  …
29           two    hatchback          fwd           front        86.6  …
30           two    hatchback          fwd           front        86.6  …
31           two    hatchback          fwd           front        93.7  …
32           two    hatchback          fwd           front        93.7  …
33           two    hatchback          fwd           front        93.7  …
34          four        sedan          fwd           front        96.5  …
35          four        wagon          fwd           front        96.5  …
36           two    hatchback          fwd           front        96.5  …
37           two    hatchback          fwd           front        96.5  …
38          four        sedan          fwd           front        96.5  …
39          four        sedan          fwd           front        96.5  …
40          four        sedan          fwd           front        96.5  …
41           two        sedan          fwd           front        96.5  …
42          four        sedan          rwd           front        94.3  …
43           two    hatchback          rwd           front        96.0  …
44          four        sedan          rwd           front       113.0  …
45          four        sedan          rwd           front       113.0  …
46           two        sedan          rwd           front       102.0  …
47           two    hatchback          fwd           front        93.1  …
48           two    hatchback          fwd           front        93.1  …
49           two    hatchback          fwd           front        93.1  …

    engine-size  fuel-system  bore  stroke compression-ratio horsepower  \
0           130         mpfi  3.47    2.68              9.00        111
1           130         mpfi  3.47    2.68              9.00        111
2           152         mpfi  2.68    3.47              9.00        154
3           109         mpfi  3.19    3.40             10.00        102
4           136         mpfi  3.19    3.40              8.00        115
5           136         mpfi  3.19    3.40              8.50        110
6           136         mpfi  3.19    3.40              8.50        110
7           136         mpfi  3.19    3.40              8.50        110
8           131         mpfi  3.13    3.40              8.30        140
9           108         mpfi  3.50    2.80              8.80        101
10          108         mpfi  3.50    2.80              8.80        101
11          164         mpfi  3.31    3.19              9.00        121
12          164         mpfi  3.31    3.19              9.00        121
13          164         mpfi  3.31    3.19              9.00        121
14          209         mpfi  3.62    3.39              8.00        182
15          209         mpfi  3.62    3.39              8.00        182
16          209         mpfi  3.62    3.39              8.00        182
17           61         2bbl  2.91    3.03              9.50         48
18           90         2bbl  3.03    3.11              9.60         70
19           90         2bbl  3.03    3.11              9.60         70
20           90         2bbl  2.97    3.23              9.41         68
```

21	90	2bbl	2.97	3.23		9.40	68
22	98	mpfi	3.03	3.39		7.60	102
23	90	2bbl	2.97	3.23		9.40	68
24	90	2bbl	2.97	3.23		9.40	68
25	90	2bbl	2.97	3.23		9.40	68
26	98	mpfi	3.03	3.39		7.60	102
27	122	2bbl	3.34	3.46		8.50	88
28	156	mfi	3.60	3.90		7.00	145
29	92	1bbl	2.91	3.41		9.60	58
30	92	1bbl	2.91	3.41		9.20	76
31	79	1bbl	2.91	3.07		10.10	60
32	92	1bbl	2.91	3.41		9.20	76
33	92	1bbl	2.91	3.41		9.20	76
34	92	1bbl	2.91	3.41		9.20	76
35	92	1bbl	2.92	3.41		9.20	76
36	110	1bbl	3.15	3.58		9.00	86
37	110	1bbl	3.15	3.58		9.00	86
38	110	1bbl	3.15	3.58		9.00	86
39	110	1bbl	3.15	3.58		9.00	86
40	110	mpfi	3.15	3.58		9.00	101
41	110	2bbl	3.15	3.58		9.10	100
42	111	2bbl	3.31	3.23		8.50	78
43	119	spfi	3.43	3.23		9.20	90
44	258	mpfi	3.63	4.17		8.10	176
45	258	mpfi	3.63	4.17		8.10	176
46	326	mpfi	3.54	2.76		11.50	262
47	91	2bbl	3.03	3.15		9.00	68
48	91	2bbl	3.03	3.15		9.00	68
49	91	2bbl	3.03	3.15		9.00	68

	peak-rpm	city-L/100km	highway-mpg	price
0	5000	11.190476	27	13495
1	5000	11.190476	27	16500
2	5000	12.368421	26	16500
3	5500	9.791667	30	13950
4	5500	13.055556	22	17450
5	5500	12.368421	25	15250
6	5500	12.368421	25	17710
7	5500	12.368421	25	18920
8	5500	13.823529	20	23875
9	5800	10.217391	29	16430
10	5800	10.217391	29	16925
11	4250	11.190476	28	20970
12	4250	11.190476	28	21105
13	4250	11.750000	25	24565
14	5400	14.687500	22	30760
15	5400	14.687500	22	41315
16	5400	15.666667	20	36880

```
17      5100     5.000000         53    5151
18      5400     6.184211         43    6295
19      5400     6.184211         43    6575
20      5500     6.351351         41    5572
21      5500     7.580645         38    6377
22      5500     9.791667         30    7957
23      5500     7.580645         38    6229
24      5500     7.580645         38    6692
25      5500     7.580645         38    7609
26      5500     9.791667         30    8558
27      5000     9.791667         30    8921
28      5000    12.368421         24   12964
29      4800     4.795918         54    6479
30      6000     7.580645         38    6855
31      5500     6.184211         42    5399
32      6000     7.833333         34    6529
33      6000     7.833333         34    7129
34      6000     7.833333         34    7295
35      6000     7.833333         34    7295
36      5800     8.703704         33    7895
37      5800     8.703704         33    9095
38      5800     8.703704         33    8845
39      5800     8.703704         33   10295
40      5800     9.791667         28   12945
41      5500     9.400000         31   10345
42      4800     9.791667         29    6785
43      5000     9.791667         29   11048
44      4750    15.666667         19   32250
45      4750    15.666667         19   35550
46      5000    18.076923         17   36000
47      5000     7.833333         31    5195
48      5000     7.580645         38    6095
49      5000     7.580645         38    6795

[50 rows x 26 columns]
```

[163]:
```python
# Save the updated DataFrame to the original CSV file
df.to_csv(r'C:\Users\dj1975\Documents\LinearRegres.csv', index=False)
```

[1]:
```python
import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv(r'C:\Users\dj1975\Documents\LinearRegres.csv')

# Verify the DataFrame is loaded
print(df.head())  # Show the first 5 rows to confirm the data
```

```
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
```

```
0        "3              122.0  alfa-romero      gas        std           two
1        "3              122.0  alfa-romero      gas        std           two
2        "1              122.0  alfa-romero      gas        std           two
3        "2              164.0         audi      gas        std          four
4        "2              164.0         audi      gas        std          four

    body-style drive-wheels engine-location  wheel-base  …  engine-size  \
0  convertible          rwd           front        88.6  …          130
1  convertible          rwd           front        88.6  …          130
2    hatchback          rwd           front        94.5  …          152
3        sedan          fwd           front        99.8  …          109
4        sedan          4wd           front        99.4  …          136

   fuel-system  bore  stroke compression-ratio horsepower  peak-rpm  \
0         mpfi  3.47    2.68               9.0        111      5000
1         mpfi  3.47    2.68               9.0        111      5000
2         mpfi  2.68    3.47               9.0        154      5000
3         mpfi  3.19    3.40              10.0        102      5500
4         mpfi  3.19    3.40               8.0        115      5500

  city-L/100km highway-mpg  price
0    11.190476          27  13495
1    11.190476          27  16500
2    12.368421          26  16500
3     9.791667          30  13950
4    13.055556          22  17450

[5 rows x 26 columns]
```

DATA NORMALIZATION: YOU ONLY NEED (1) BUT HERE ARE (3) TECHNIQUES

[3]:
```python
#Simple Feature Scaling
df["length"] = df["length"]/df["length"].max()
```

[5]:
```python
#Inspect the new length column by printing its statistics
print(df["length"].describe())
```

```
count    201.000000
mean       0.837102
std        0.059213
min        0.678039
25%        0.801538
50%        0.832292
75%        0.881788
max        1.000000
Name: length, dtype: float64
```

[9]:
```python
# Apply Min-Max normalization
df["length"] = (df["length"] - df["length"].min()) / (df["length"].max() -␣
↪df["length"].min())
```

[11]:
```python
# Check the min and max values of the normalized column
print("Minimum value:", df["length"].min())
print("Maximum value:", df["length"].max())
```

```
Minimum value: 0.0
Maximum value: 1.0
```

[13]:
```python
#Apply Z-Score Normalizaton
df["length"] = (df["length"] - df["length"].mean()) / df["length"].std()
```

[15]:
```python
# Check the new mean and standard deviation of the "length" column
print("Mean:", df["length"].mean())
print("Standard Deviation:", df["length"].std())
```

```
Mean: 2.1210230918211945e-16
Standard Deviation: 0.9999999999999997
```

[19]:
```python
#Save the updates
df.to_csv(r"C:\Users\dj1975\Documents\LinearRegres.csv", index=False)
```

[ ]:
```
BINNING In PYTHON
```

[22]:
```python
import numpy as np
import pandas as pd

# Create bins for the "price" column
bins = np.linspace(min(df["price"]), max(df["price"]), 4)
```

[24]:
```python
# Group names for the bins
group_names = ["Low", "Medium", "High"]
```

[26]:
```python
# Apply binning to the "price" column
df["price-binned"] = pd.cut(df["price"], bins, labels=group_names,␣
↪include_lowest=True)
```

[29]:
```python
print(df.head())
```

```
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3              122.0  alfa-romero       gas        std          two
1        "3              122.0  alfa-romero       gas        std          two
2        "1              122.0  alfa-romero       gas        std          two
3        "2              164.0         audi       gas        std         four
4        "2              164.0         audi       gas        std         four
```

```
   body-style drive-wheels engine-location  wheel-base  …  fuel-system  \
0  convertible          rwd           front        88.6  …         mpfi
1  convertible          rwd           front        88.6  …         mpfi
2    hatchback          rwd           front        94.5  …         mpfi
3        sedan          fwd           front        99.8  …         mpfi
4        sedan          4wd           front        99.4  …         mpfi

   bore  stroke  compression-ratio horsepower peak-rpm  city-L/100km  \
0  3.47    2.68                9.0        111     5000     11.190476
1  3.47    2.68                9.0        111     5000     11.190476
2  2.68    3.47                9.0        154     5000     12.368421
3  3.19    3.40               10.0        102     5500      9.791667
4  3.19    3.40                8.0        115     5500     13.055556

  highway-mpg  price price-binned
0          27  13495          Low
1          27  16500          Low
2          26  16500          Low
3          30  13950          Low
4          22  17450          Low

[5 rows x 27 columns]
```

[31]:
```
#Print Bins
print(bins)
```

```
[ 5118.         18545.33333333 31972.66666667 45400.        ]
```

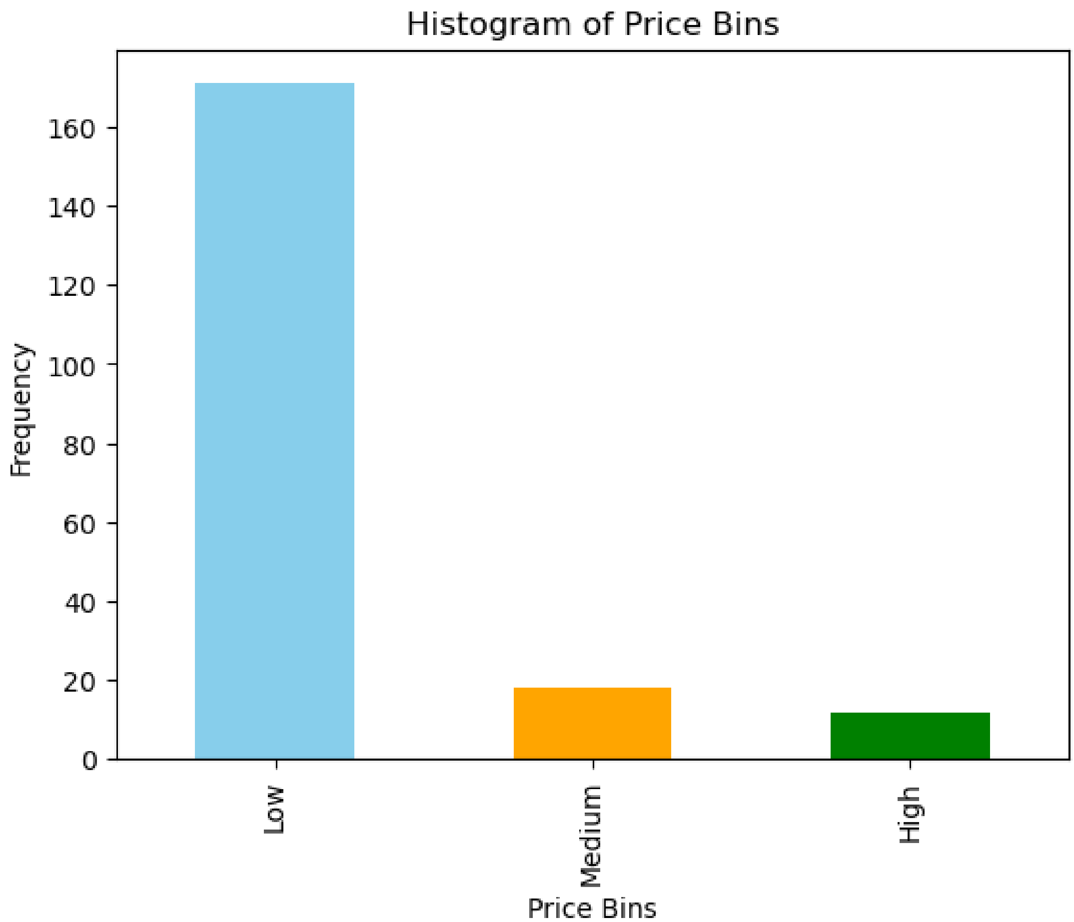
[33]:
```
#Confirms that the price values into the specified bins based on the price␣
↪range.
df['price-binned'].value_counts()
```

[33]:
```
price-binned
Low       171
Medium     18
High       12
Name: count, dtype: int64
```

[ ]:
```
CREATE A HISTOGRAM Of THE BINS
```

[35]:
```
import matplotlib.pyplot as plt

# Plot the histogram of the 'price-binned' column
df['price-binned'].value_counts().plot(kind='bar', color=['skyblue', 'orange',␣
↪'green'])

# Adding labels and title
```

```
plt.xlabel('Price Bins')
plt.ylabel('Frequency')
plt.title('Histogram of Price Bins')

# Show the plot
plt.show()
```



GET DUMMIES METHOD: TURN CATEGORICAL VARIABLES INTO QUANTITATIVE VARIABLES

[45]:
```
# Apply one-hot encoding to the 'fuel-type' column
df_encoded = pd.get_dummies(df, columns=["fuel-type"], drop_first=False)
df_encoded[["fuel-type_diesel", "fuel-type_gas"]] =␣
↪df_encoded[["fuel-type_diesel", "fuel-type_gas"]].astype(int)



# Display the first few rows to see the result
#You should see 0 or 1 in "fuel-type_diesel" or "fuel_type_gas"
```

```
print(df_encoded.head())
```

```
  symboling  normalized-losses         make aspiration num-of-doors  \
0        "3              122.0  alfa-romero        std          two
1        "3              122.0  alfa-romero        std          two
2        "1              122.0  alfa-romero        std          two
3        "2              164.0         audi        std         four
4        "2              164.0         audi        std         four

    body-style drive-wheels engine-location  wheel-base    length  …  \
0  convertible          rwd           front        88.6 -0.438315  …
1  convertible          rwd           front        88.6 -0.438315  …
2    hatchback          rwd           front        94.5 -0.243544  …
3        sedan          fwd           front        99.8  0.194690  …
4        sedan          4wd           front        99.4  0.194690  …

   stroke  compression-ratio  horsepower peak-rpm city-L/100km  highway-mpg  \
0    2.68                9.0         111     5000    11.190476           27
1    2.68                9.0         111     5000    11.190476           27
2    3.47                9.0         154     5000    12.368421           26
3    3.40               10.0         102     5500     9.791667           30
4    3.40                8.0         115     5500    13.055556           22

   price price-binned fuel-type_diesel  fuel-type_gas
0  13495          Low                0              1
1  16500          Low                0              1
2  16500          Low                0              1
3  13950          Low                0              1
4  17450          Low                0              1

[5 rows x 28 columns]
```

[48]:
```
#Save the updates
df.to_csv(r"C:\Users\dj1975\Documents\LinearRegres.csv", index=False)
```

[52]:
```
# Apply one-hot encoding to the 'fuel-type' column
df_encoded = pd.get_dummies(df, columns=["fuel-type"], drop_first=False)
df_encoded[["fuel-type_diesel", "fuel-type_gas"]] =␣
 ↪df_encoded[["fuel-type_diesel", "fuel-type_gas"]].astype(int)


# Display the first few rows to see the result
#You should see 0 or 1 in "fuel-type_diesel" or "fuel_type_gas"
print(df_encoded.head())
```

```
  symboling  normalized-losses         make aspiration num-of-doors  \
0        "3              122.0  alfa-romero        std          two
```

```
1       "3             122.0  alfa-romero       std          two
2       "1             122.0  alfa-romero       std          two
3       "2             164.0         audi       std         four
4       "2             164.0         audi       std         four

    body-style drive-wheels engine-location  wheel-base    length  …  \
0  convertible          rwd           front        88.6 -0.438315  …
1  convertible          rwd           front        88.6 -0.438315  …
2    hatchback          rwd           front        94.5 -0.243544  …
3        sedan          fwd           front        99.8  0.194690  …
4        sedan          4wd           front        99.4  0.194690  …

   stroke  compression-ratio  horsepower peak-rpm city-L/100km  highway-mpg  \
0    2.68                9.0         111     5000    11.190476           27
1    2.68                9.0         111     5000    11.190476           27
2    3.47                9.0         154     5000    12.368421           26
3    3.40               10.0         102     5500     9.791667           30
4    3.40                8.0         115     5500    13.055556           22

   price price-binned fuel-type_diesel  fuel-type_gas
0  13495          Low                0              1
1  16500          Low                0              1
2  16500          Low                0              1
3  13950          Low                0              1
4  17450          Low                0              1

[5 rows x 28 columns]
```

[54]:
```
#Save the updates
df.to_csv(r"C:\Users\dj1975\Documents\LinearRegres.csv", index=False)
```

[56]:
```
# Count the occurrences of 'gas' and 'diesel' in the 'fuel-type' column
fuel_count = df['fuel-type'].value_counts()
print(fuel_count)
```

```
fuel-type
gas       181
diesel     20
Name: count, dtype: int64
```

MODULE 3: EXPLORATORY DATA ANALYSIS

[3]:
```
import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv(r'C:\Users\dj1975\Documents\LinearRegres.csv')

# Verify the DataFrame is loaded
```

```python
print(df.head())  # Show the first 5 rows to confirm the data
```

```
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3              122.0  alfa-romero       gas        std          two
1        "3              122.0  alfa-romero       gas        std          two
2        "1              122.0  alfa-romero       gas        std          two
3        "2              164.0         audi       gas        std         four
4        "2              164.0         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  fuel-system  \
0  convertible          rwd           front        88.6  …         mpfi
1  convertible          rwd           front        88.6  …         mpfi
2    hatchback          rwd           front        94.5  …         mpfi
3        sedan          fwd           front        99.8  …         mpfi
4        sedan          4wd           front        99.4  …         mpfi

   bore  stroke  compression-ratio horsepower peak-rpm  city-L/100km  \
0  3.47    2.68                9.0        111     5000     11.190476
1  3.47    2.68                9.0        111     5000     11.190476
2  2.68    3.47                9.0        154     5000     12.368421
3  3.19    3.40               10.0        102     5500      9.791667
4  3.19    3.40                8.0        115     5500     13.055556

  highway-mpg  price price-binned
0          27  13495          Low
1          27  16500          Low
2          26  16500          Low
3          30  13950          Low
4          22  17450          Low

[5 rows x 27 columns]
```

[5]:
```python
df.describe()
```

[5]:
```
       normalized-losses  wheel-base        length       width      height  \
count          201.00000  201.000000  2.010000e+02  201.000000  201.000000
mean           122.00000   98.797015  2.121023e-16   65.889055   53.766667
std             31.99625    6.066366  1.000000e+00    2.101471    2.447822
min             65.00000   86.600000 -2.686295e+00   60.300000   47.800000
25%            101.00000   94.500000 -6.006241e-01   64.100000   52.000000
50%            122.00000   97.000000 -8.123525e-02   65.500000   54.100000
75%            137.00000  102.400000  7.546561e-01   66.600000   55.500000
max            256.00000  120.900000  2.751057e+00   72.000000   59.800000

       curb-weight  engine-size  compression-ratio  city-L/100km  highway-mpg  \
count   201.000000   201.000000         201.000000    201.000000   201.000000
mean   2555.666667   126.875622          10.164279      9.944145    30.686567
```

```
std      517.296727    41.546834             4.004965       2.534599     6.815150
min     1488.000000    61.000000             7.000000       4.795918    16.000000
25%     2169.000000    98.000000             8.600000       7.833333    25.000000
50%     2414.000000   120.000000             9.000000       9.791667    30.000000
75%     2926.000000   141.000000             9.400000      12.368421    34.000000
max     4066.000000   326.000000            23.000000      18.076923    54.000000

              price
count    201.000000
mean   13207.129353
std     7947.066342
min     5118.000000
25%     7775.000000
50%    10295.000000
75%    16500.000000
max    45400.000000
```

SUMMARIZE CATEGORICAL DATA USING VALUE_COUNTS() METHOD

[9]:
```
# Summarize the categorical data
drive_wheels_counts = df["drive-wheels"].value_counts()
```

[11]:
```
# Convert the Series to a DataFrame for better readability
drive_wheels_counts_df = drive_wheels_counts.reset_index()
```

[13]:
```
# Change the column names to make things easier to read
drive_wheels_counts_df.columns = ['drive_wheels', 'value_counts']
```

[15]:
```
# Set the 'drive_wheels' column as the index for better formatting (optional)
drive_wheels_counts_df.set_index('drive_wheels', inplace=True)
```

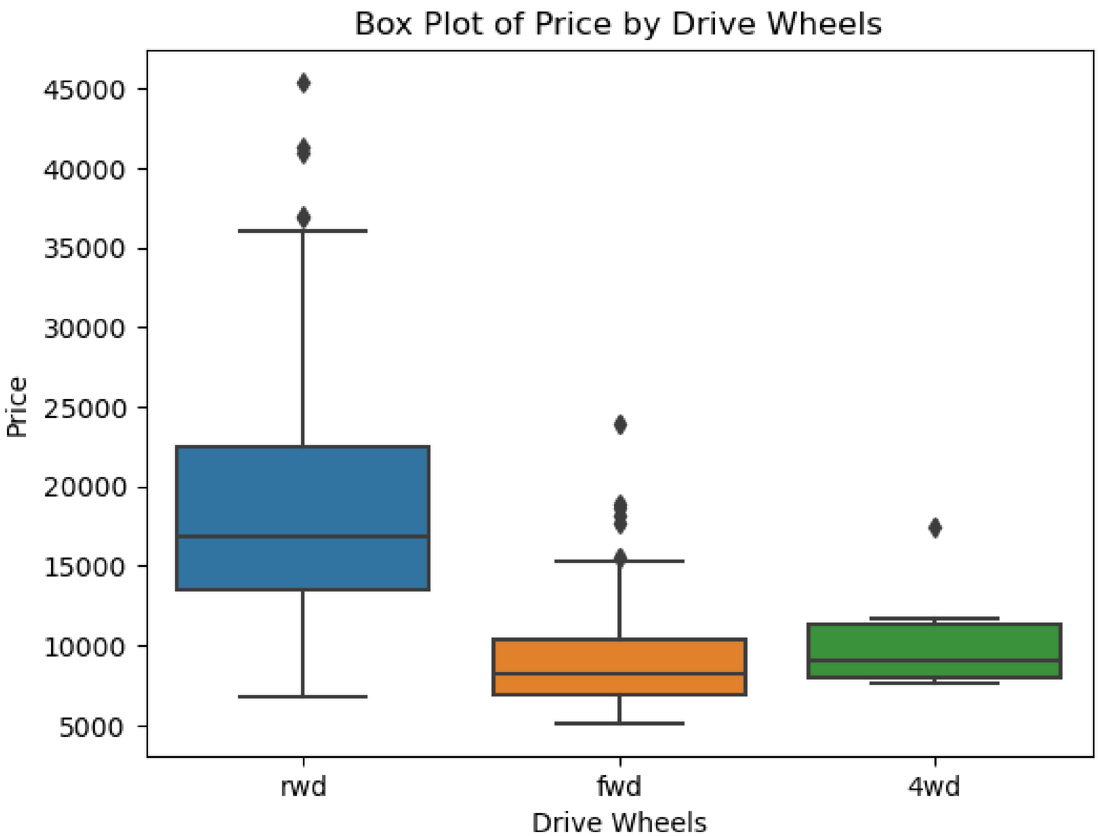
[17]:
```
# Display the result
print(drive_wheels_counts_df)
```

```
              value_counts
drive_wheels
fwd                    118
rwd                     75
4wd                      8
```

Creating A Box Plot: Analyze “drive-wheels” relationship to “price”

[19]:
```
# Import libraries
import seaborn as sns
import matplotlib.pyplot as plt

# Create a box plot
sns.boxplot(x="drive-wheels", y="price", data=df)
```

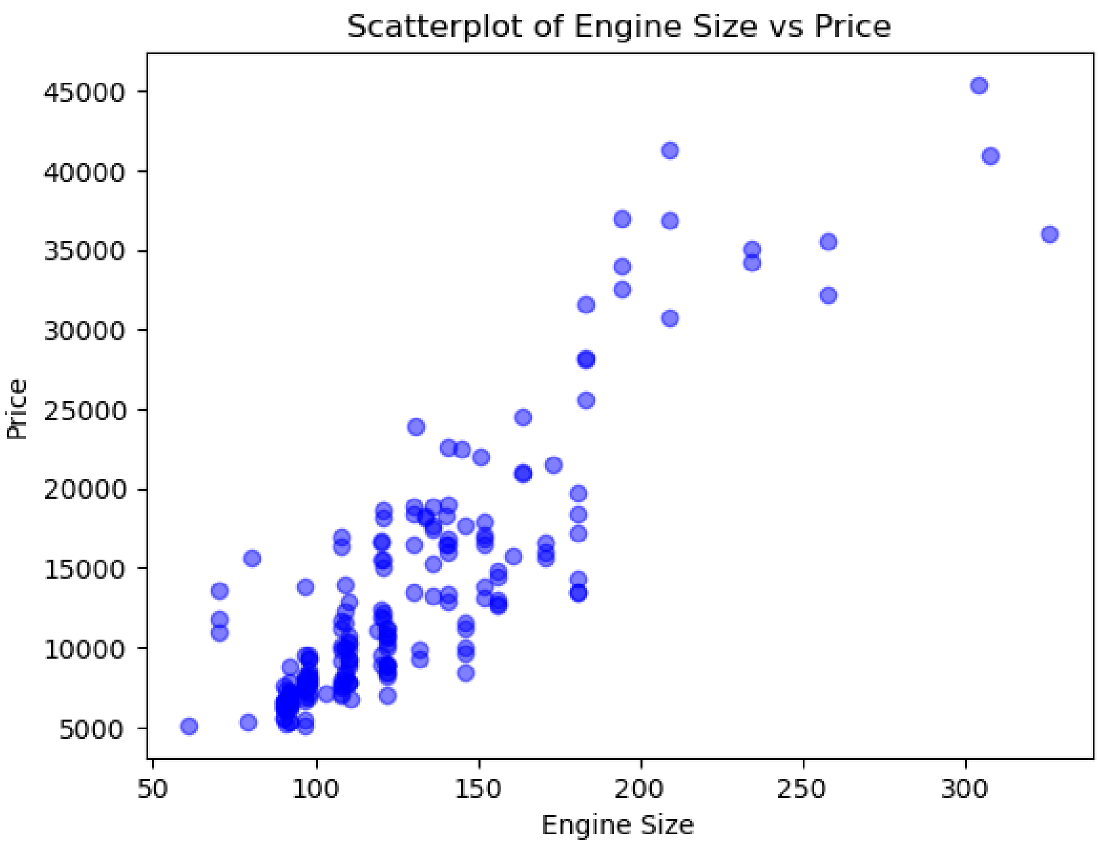
```
# Customize plot
plt.title("Box Plot of Price by Drive Wheels")
plt.xlabel("Drive Wheels")
plt.ylabel("Price")

# Show plot
plt.show()
```



Creating A Scatter Plot: Understand The Relationship Between "engine-size" and "price"

[21]:
```
import matplotlib.pyplot as plt

# Define x and y
x = df["engine-size"]
y = df["price"]

# Create scatter plot
plt.scatter(x, y, color='blue', alpha=0.5)
```

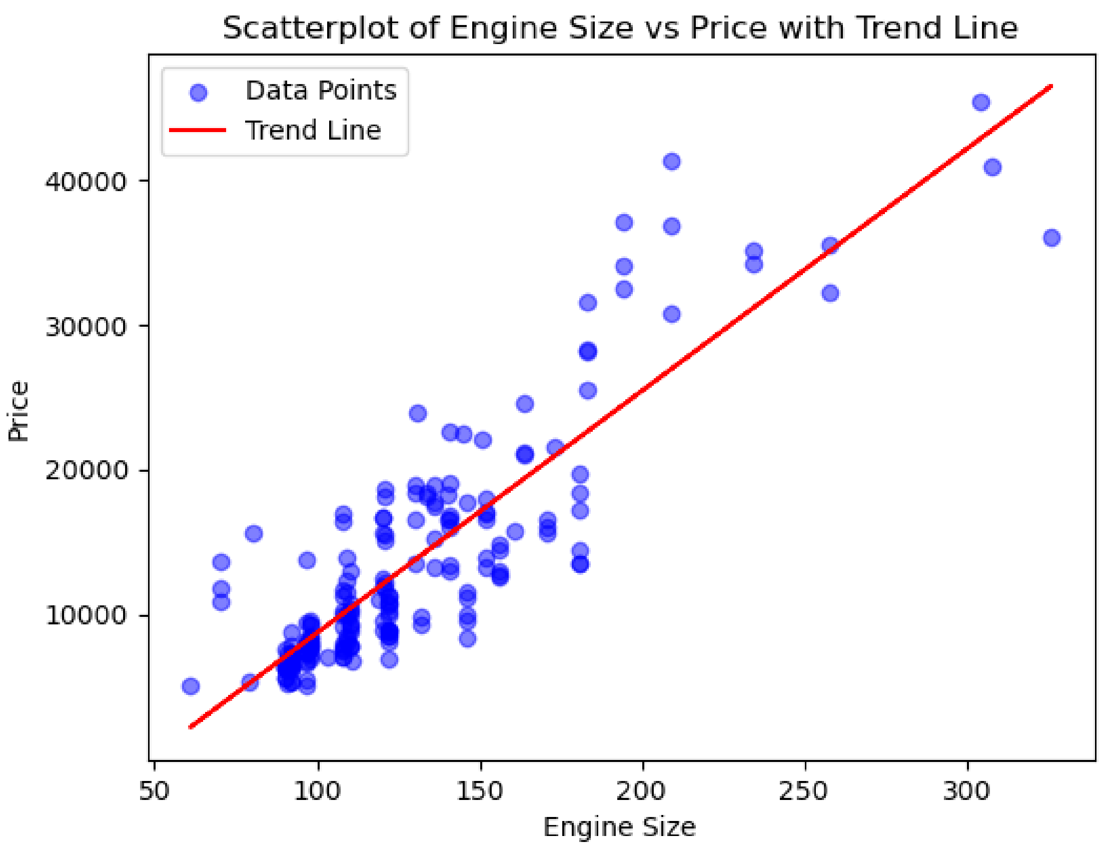
```
# Add title and labels
plt.title("Scatterplot of Engine Size vs Price")
plt.xlabel("Engine Size")
plt.ylabel("Price")

# Display plot
plt.show()
```



[23]:

```
#Scatterplot With Trend Line
import matplotlib.pyplot as plt
import numpy as np

# Define x and y
x = df["engine-size"]
y = df["price"]

# Scatter plot
plt.scatter(x, y, color='blue', alpha=0.5, label="Data Points")

# Calculate trend line (linear regression)
```

```python
coefficients = np.polyfit(x, y, 1)  # Degree 1 for linear
trendline = np.poly1d(coefficients)

# Add trend line to plot
plt.plot(x, trendline(x), color='red', label="Trend Line")

# Add title and labels
plt.title("Scatterplot of Engine Size vs Price with Trend Line")
plt.xlabel("Engine Size")
plt.ylabel("Price")

# Add legend
plt.legend()

# Show plot
plt.show()
```



GROUPING DATA: IS THERE ANY RELATIONSHIP BETWEEN DIFFERENT “drive system” and “price” of vehicles?

[25]:
```python
import pandas as pd

# Assuming 'df' is already defined and has the columns 'drive-wheels',␣
↪'body-style', and 'price'

# Pick out the columns you are analyzing
dftest = df[["drive-wheels", "body-style", "price"]]

# Group the data by 'drive-wheels' and 'body-style' and calculate the mean␣
↪price for each group
df_grp = dftest.groupby(["drive-wheels", "body-style"], as_index=False).mean()

# Display the grouped data
print(df_grp)
```

```
   drive-wheels   body-style         price
0           4wd    hatchback   7603.000000
1           4wd        sedan  12647.333333
2           4wd        wagon   9095.750000
3           fwd  convertible  11595.000000
4           fwd      hardtop   8249.000000
5           fwd    hatchback   8396.387755
6           fwd        sedan   9811.800000
7           fwd        wagon   9997.333333
8           rwd  convertible  23949.600000
9           rwd      hardtop  24202.714286
10          rwd    hatchback  14337.777778
11          rwd        sedan  21711.833333
12          rwd        wagon  16994.222222
```

Transform This Data To A Pivot Table Making It Easier To Read

[29]:
```python
# Assuming df_grp already has the grouped data with 'drive-wheels',␣
↪'body-style', and mean 'price'

# Pivot the data for better visualization
df_pivot = df_grp.pivot(index="drive-wheels", columns="body-style",␣
↪values="price")

# Display the pivoted DataFrame
print(df_pivot)
```

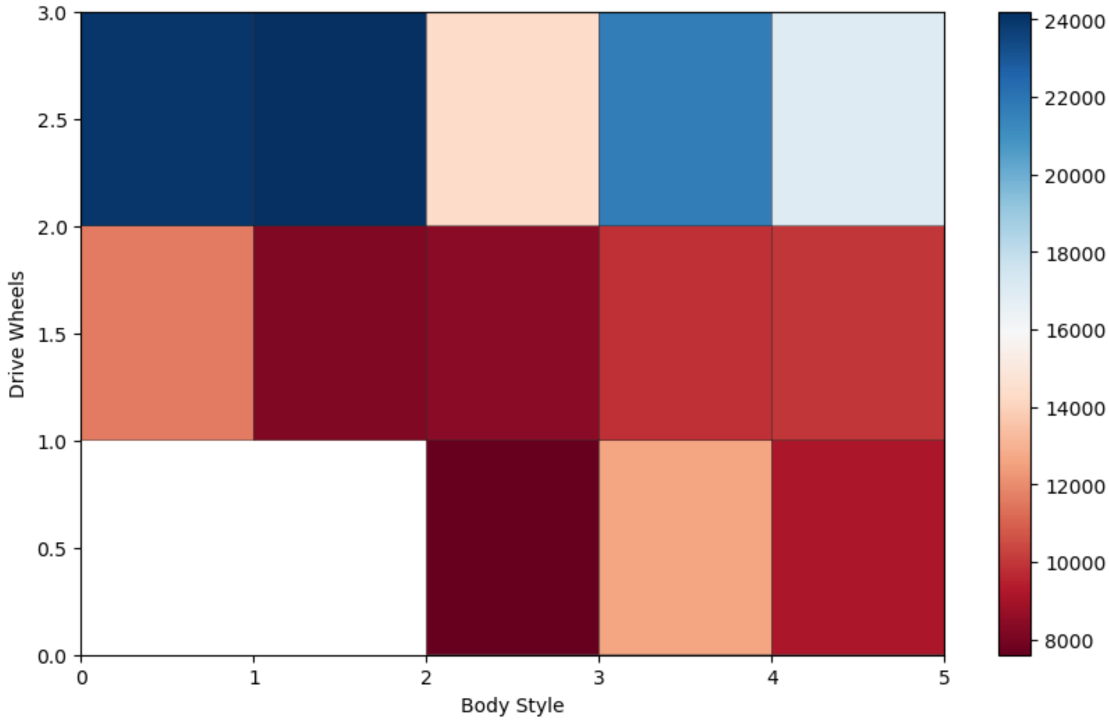
```
body-style    convertible       hardtop     hatchback         sedan  \
drive-wheels
4wd                   NaN           NaN   7603.000000  12647.333333
fwd               11595.0   8249.000000   8396.387755   9811.800000
rwd               23949.6  24202.714286  14337.777778  21711.833333
```

```
body-style           wagon
drive-wheels
4wd              9095.750000
fwd              9997.333333
rwd             16994.222222
```

Create A Heatmap: Plot Target Variable Against Multiple Variables

[31]:
```
import matplotlib.pyplot as plt

# Assuming df_pivot is your pivoted DataFrame

# Create a heatmap using pcolor
plt.figure(figsize=(10, 6))  # Optional: Adjust the figure size
plt.pcolor(df_pivot, cmap="RdBu", edgecolors='k')  # Optional: Add grid lines
plt.colorbar()  # Add a color bar to the side for reference

# Set labels for the axes
plt.xlabel('Body Style')
plt.ylabel('Drive Wheels')

# Display the heatmap
plt.show()
```

```
/tmp/ipykernel_149/2500661161.py:8: MatplotlibDeprecationWarning: Getting the
array from a PolyQuadMesh will return the full array in the future
(uncompressed). To get this behavior now set the PolyQuadMesh with a 2D array
.set_array(data2d).
  plt.colorbar()  # Add a color bar to the side for reference
```

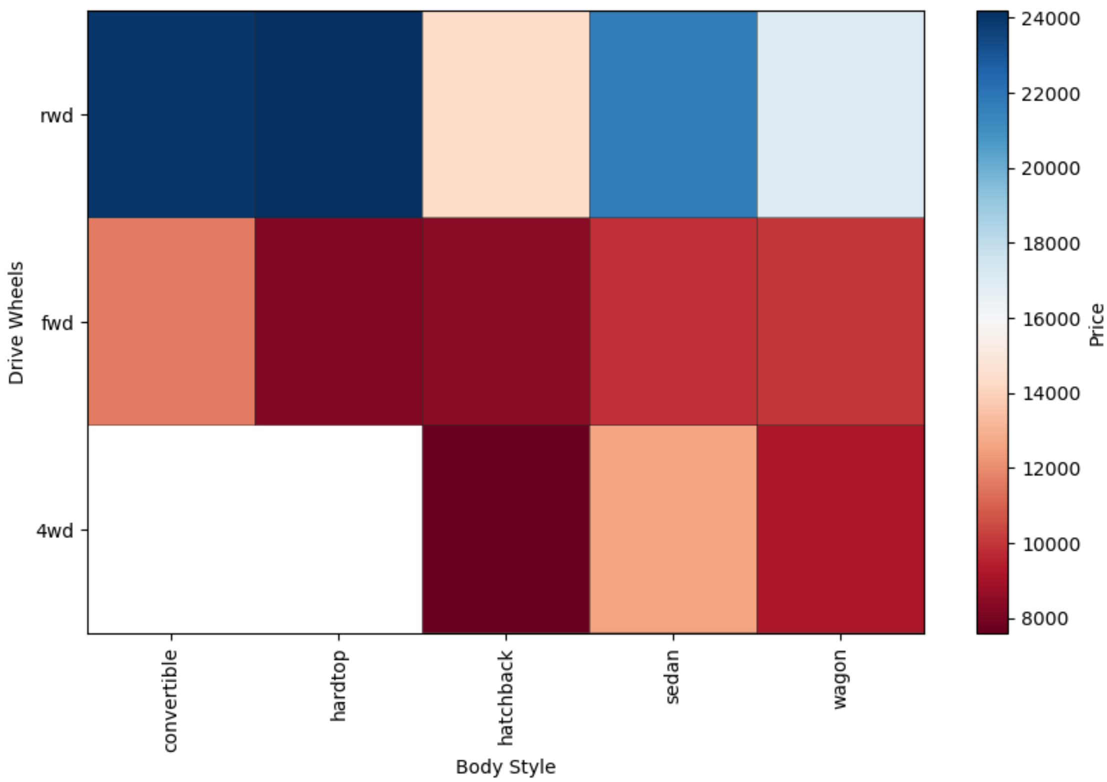


AS YOU CAN SEE ABOVE, THE HEATMAP IS NOT DESCRIPTIVE ENOUGH. THE HEATMAP BELOW IS BETTER

[35]:
```
# Create a heatmap using pcolor
plt.figure(figsize=(10, 6))  # Optional: Adjust the figure size
plt.pcolor(df_pivot, cmap="RdBu", edgecolors='k')  # Optional: Add grid lines

# Set labels for the axes
plt.xlabel('Body Style')  # X-axis: body style (convertible, sedan, etc.)
plt.ylabel('Drive Wheels')  # Y-axis: drive type (fwd, rwd, 4wd)

# Adjust tick labels for the correct categorization
plt.xticks([i + 0.5 for i in range(len(df_pivot.columns))], df_pivot.columns,
 ↪rotation=90)
plt.yticks([i + 0.5 for i in range(len(df_pivot.index))], df_pivot.index)

# Add a color bar to the side for reference and set the label to "Price"
cbar = plt.colorbar()
cbar.set_label('Price')

# Display the heatmap
plt.show()
```

```
/tmp/ipykernel_149/2620911388.py:14: MatplotlibDeprecationWarning: Getting the
```

```
array from a PolyQuadMesh will return the full array in the future
(uncompressed). To get this behavior now set the PolyQuadMesh with a 2D array
.set_array(data2d).
  cbar = plt.colorbar()
```



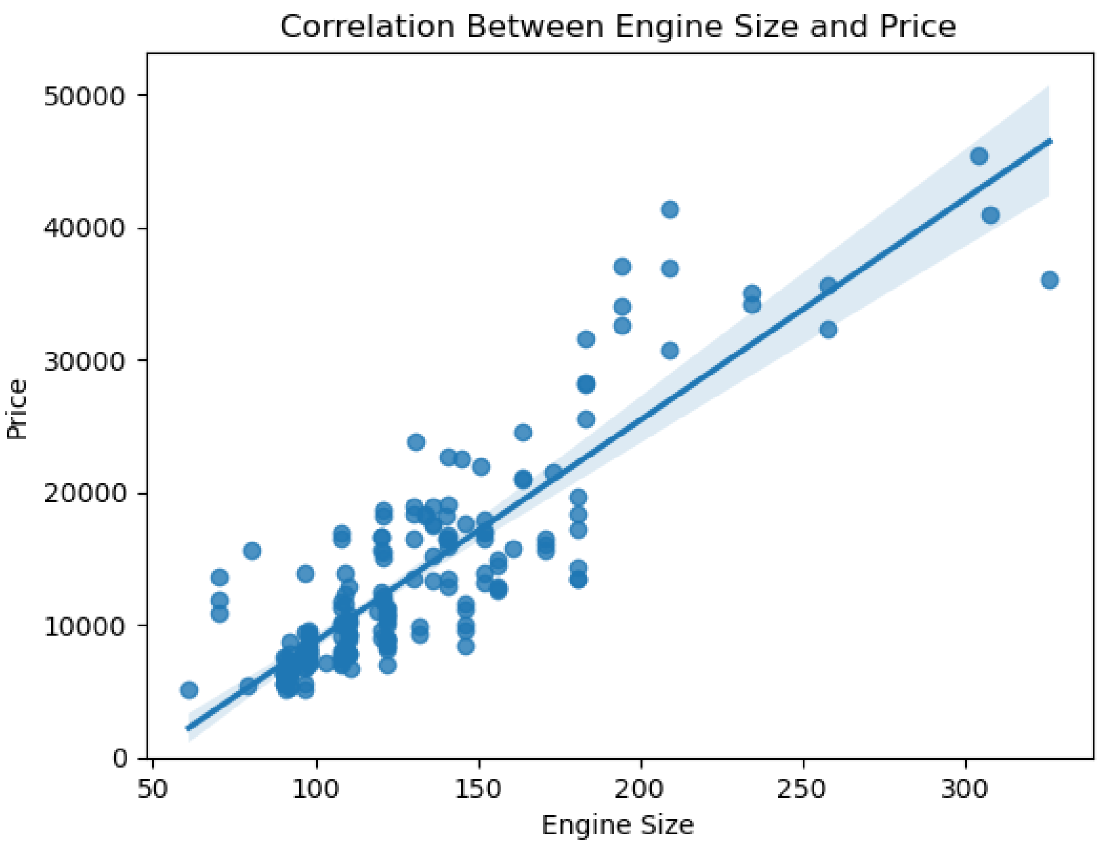
[1]:
```python
import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv(r'C:\Users\dj1975\Documents\LinearRegres.csv')

# Verify the DataFrame is loaded
print(df.head())  # Show the first 5 rows to confirm the data
```

```
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3              122.0  alfa-romero       gas        std          two
1        "3              122.0  alfa-romero       gas        std          two
2        "1              122.0  alfa-romero       gas        std          two
3        "2              164.0         audi       gas        std         four
4        "2              164.0         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  fuel-system  \
0  convertible          rwd           front        88.6  …         mpfi
1  convertible          rwd           front        88.6  …         mpfi
```

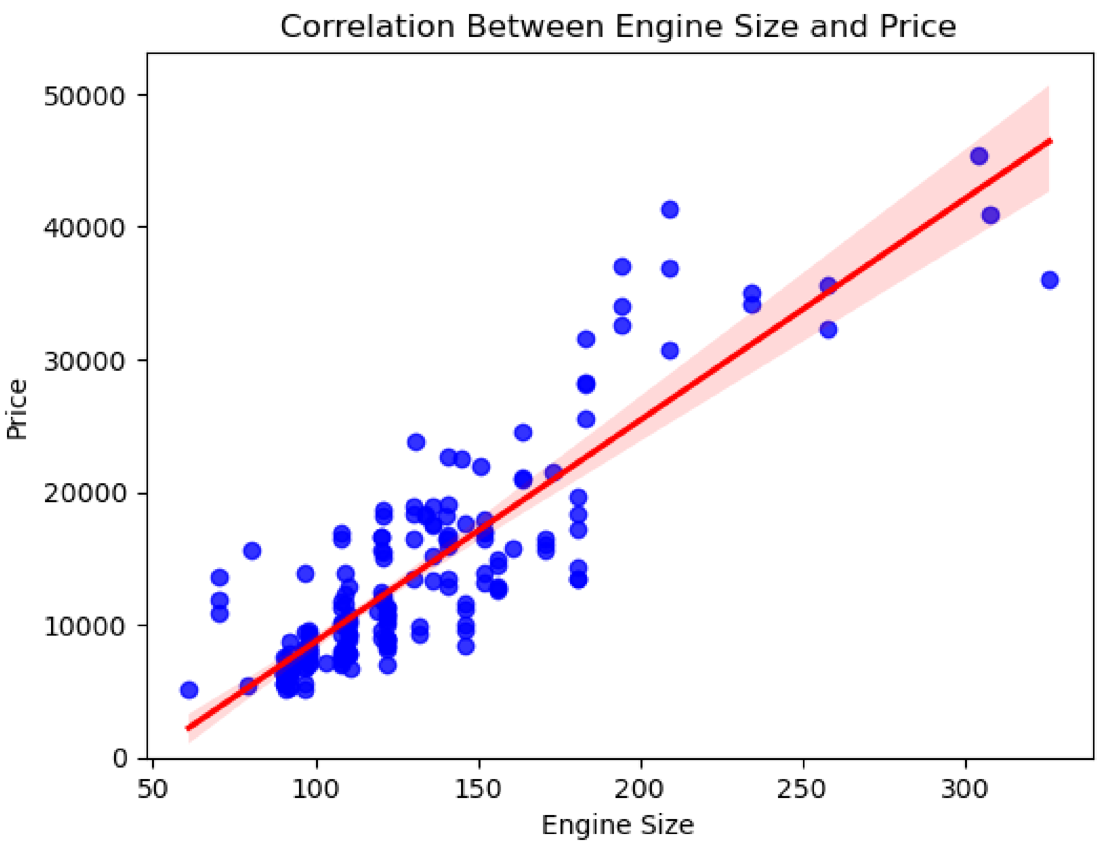
```
2    hatchback           rwd          front         94.5  …          mpfi
3          sedan         fwd          front         99.8  …          mpfi
4          sedan         4wd          front         99.4  …          mpfi

   bore  stroke  compression-ratio horsepower peak-rpm  city-L/100km  \
0  3.47    2.68                9.0         111     5000     11.190476
1  3.47    2.68                9.0         111     5000     11.190476
2  2.68    3.47                9.0         154     5000     12.368421
3  3.19    3.40               10.0         102     5500      9.791667
4  3.19    3.40                8.0         115     5500     13.055556

  highway-mpg  price price-binned
0          27  13495          Low
1          27  16500          Low
2          26  16500          Low
3          30  13950          Low
4          22  17450          Low

[5 rows x 27 columns]
```

WHAT IS CORRELATION?

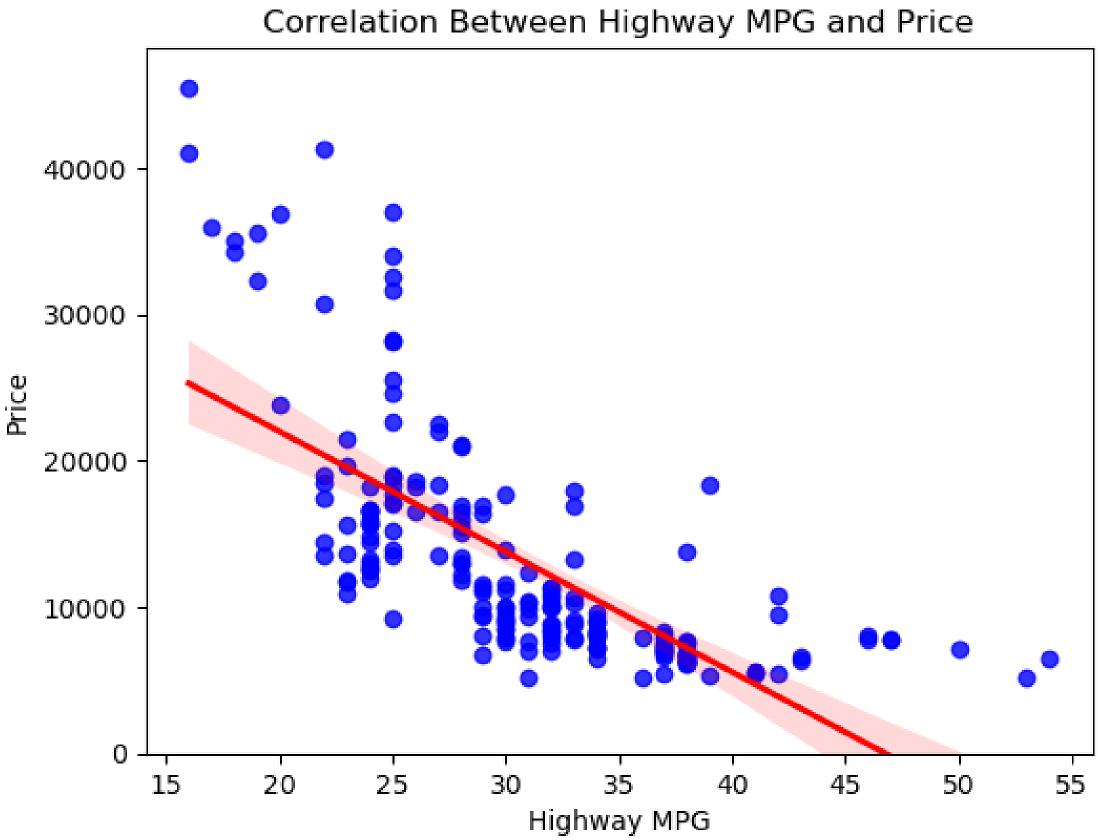
[3]:
```python
#Positive Linear Relationship
import seaborn as sns
import matplotlib.pyplot as plt

# Create a regression plot to visualize the correlation
sns.regplot(x="engine-size", y="price", data=df)

# Set the y-axis limit to start from 0 for clarity
plt.ylim(0)

# Add titles and labels for better interpretation
plt.title("Correlation Between Engine Size and Price")
plt.xlabel("Engine Size")
plt.ylabel("Price")

# Display the plot
plt.show()
```



[5]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Create a scatter plot with a trend line (regression line)
sns.regplot(x="engine-size", y="price", data=df, scatter_kws={"color": "blue"},␣
 ↪line_kws={"color": "red"})

# Set the y-axis limit to start from 0
plt.ylim(0)

# Add title and axis labels for clarity
plt.title("Correlation Between Engine Size and Price")
plt.xlabel("Engine Size")
plt.ylabel("Price")

# Display the plot
plt.show()
```



[7]:
```
#Negative Linear Relationship
import seaborn as sns
import matplotlib.pyplot as plt

# Create a scatter plot with a trend line (regression line)
sns.regplot(x="highway-mpg", y="price", data=df, scatter_kws={"color": "blue"},␣
 ↪line_kws={"color": "red"})

# Set the y-axis limit to start from 0
plt.ylim(0)

# Add title and axis labels for clarity
plt.title("Correlation Between Highway MPG and Price")
plt.xlabel("Highway MPG")
plt.ylabel("Price")

# Display the plot
plt.show()
```

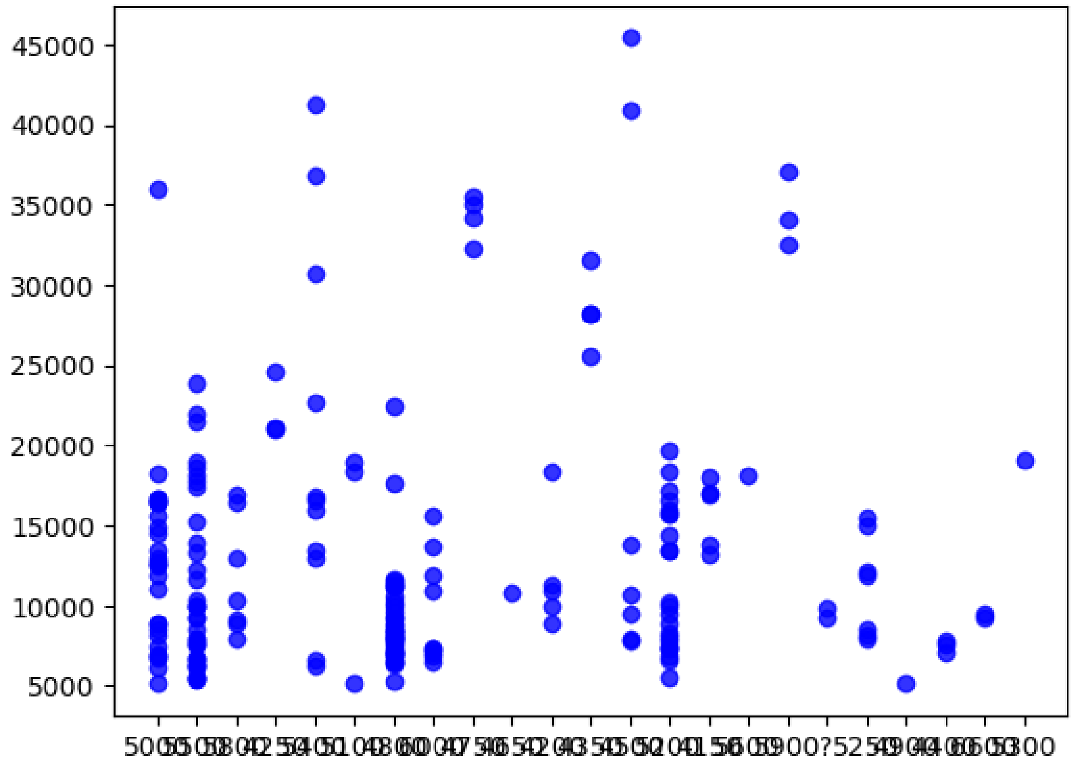


[9]:
```
#Weak Linear Correlation
import seaborn as sns
import matplotlib.pyplot as plt

# Create a scatter plot with a trend line (regression line)
sns.regplot(x="peak-rpm", y="price", data=df, scatter_kws={"color": "blue"},␣
 ↪line_kws={"color": "red"})

# Set the y-axis limit to start from 0
plt.ylim(0)

# Add title and axis labels for clarity
plt.title("Correlation Between Peak RPM and Price")
plt.xlabel("Peak RPM")
plt.ylabel("Price")

# Display the plot
plt.show()
```

```
---------------------------------------------------------------------------
```

```
UFuncTypeError                            Traceback (most recent call last)
Cell In[9], line 6
      3 import matplotlib.pyplot as plt
      5 # Create a scatter plot with a trend line (regression line)
----> 6␣
  ↪sns.regplot(x="peak-rpm", y="price", data=df, scatter_kws={"color": "blue"}, line_kws={"co
      8 # Set the y-axis limit to start from 0
      9 plt.ylim(0)

File /opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/seabor /
  ↪regression.py:759, in regplot(data, x, y, x_estimator, x_bins, x_ci, scatter,␣
  ↪fit_reg, ci, n_boot, units, seed, order, logistic, lowess, robust, logx,␣
  ↪x_partial, y_partial, truncate, dropna, x_jitter, y_jitter, label, color,␣
  ↪marker, scatter_kws, line_kws, ax)
    757 scatter_kws["marker"] = marker
    758 line_kws = {} if line_kws is None else copy.copy(line_kws)
--> 759 plotter.plot(ax, scatter_kws, line_kws)
    760 return ax

File /opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/seabor /
  ↪regression.py:368, in _RegressionPlotter.plot(self, ax, scatter_kws, line_kws
    365     self.scatterplot(ax, scatter_kws)
    367 if self.fit_reg:
--> 368     self.lineplot(ax, line_kws)
    370 # Label the axes
    371 if hasattr(self.x, "name"):

File /opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/seabor /
  ↪regression.py:413, in _RegressionPlotter.lineplot(self, ax, kws)
    411 """Draw the model."""
    412 # Fit the regression model
--> 413 grid, yhat, err_bands = self.fit_regression(ax)
    414 edges = grid[0], grid[-1]
    416 # Get set default aesthetics

File /opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/seabor /
  ↪regression.py:199, in _RegressionPlotter.fit_regression(self, ax, x_range,␣
  ↪grid)
    197         else:
    198             x_min, x_max = ax.get_xlim()
--> 199     grid = np.linspace(x_min, x_max, 100)
    200 ci = self.ci
    202 # Fit the regression

File /opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/numpy/
  ↪core/function_base.py:129, in linspace(start, stop, num, endpoint, retstep,␣
  ↪dtype, axis)
    125 div = (num - 1) if endpoint else num
    127 # Convert float/complex array scalars to float, gh-3504
```

```
    128 # and make sure one can use variables that have an __array_interface__,␣
  ↪gh-6634
--> 129 start = asanyarray(start) * 1.0
    130 stop  = asanyarray(stop)  * 1.0
    132 dt = result_type(start, stop, float(num))

UFuncTypeError: ufunc 'multiply' did not contain a loop with signature matching␣
  ↪types (dtype('<U4'), dtype('float64')) -> None
```



The error above indicates that the peak-rpm or price column in your dataset contains non-numeric data, such as strings.

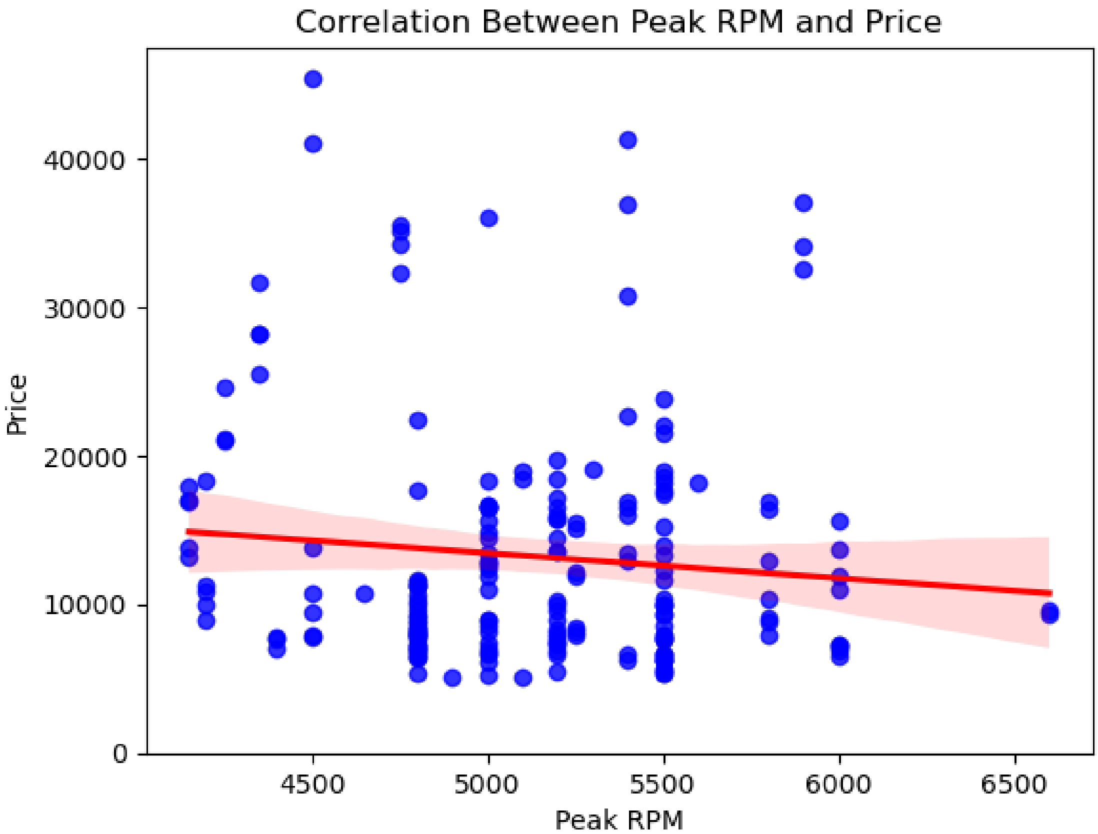
[11]:
```
#Inspect the Data: Check for non-numeric entries in the peak-rpm and price␣
  ↪columns.
print(df["peak-rpm"].unique())
print(df["price"].unique())
```

```
['5000' '5500' '5800' '4250' '5400' '5100' '4800' '6000' '4750' '4650'
 '4200' '4350' '4500' '5200' '4150' '5600' '5900' '?' '5250' '4900' '4400'
 '6600' '5300']
[13495 16500 13950 17450 15250 17710 18920 23875 16430 16925 20970 21105
 24565 30760 41315 36880  5151  6295  6575  5572  6377  7957  6229  6692
```

```
 7609  8558  8921 12964  6479  6855  5399  6529  7129  7295  7895  9095
 8845 10295 12945 10345  6785 11048 32250 35550 36000  5195  6095  6795
 6695  7395 10945 11845 13645 15645  8495 10595 10245 10795 11245 18280
18344 25552 28248 28176 31600 34184 35056 40960 45400 16503  5389  6189
 6669  7689  9959  8499 12629 14869 14489  6989  8189  9279  5499  7099
 6649  6849  7349  7299  7799  7499  7999  8249  8949  9549 13499 14399
17199 19699 18399 11900 13200 12440 13860 15580 16900 16695 17075 16630
17950 18150 12764 22018 32528 34028 37028  9295  9895 11850 12170 15040
15510 18620  5118  7053  7603  7126  7775  9960  9233 11259  7463 10198
 8013 11694  5348  6338  6488  6918  7898  8778  6938  7198  7788  7738
 8358  9258  8058  8238  9298  9538  8449  9639  9989 11199 11549 17669
 8948 10698  9988 10898 11248 16558 15998 15690 15750  7975  7995  8195
 9495  9995 11595  9980 13295 13845 12290 12940 13415 15985 16515 18420
18950 16845 19045 21485 22470 22625]
```

[13]:
```python
#Handle Missing or Non-Numeric Values
#Replace non-numeric values with NaN and convert the column to numeric
#Drop or impute the rows with missing or non-numeric values
import pandas as pd
import numpy as np

# Replace non-numeric values with NaN and convert to numeric
df["peak-rpm"] = pd.to_numeric(df["peak-rpm"], errors="coerce")
df["price"] = pd.to_numeric(df["price"], errors="coerce")

# Drop rows with NaN values in the specified columns
df = df.dropna(subset=["peak-rpm", "price"])
```

[17]:
```python
#Weak Correlation
import seaborn as sns
import matplotlib.pyplot as plt

# Create a scatter plot with a trend line
sns.regplot(x="peak-rpm", y="price", data=df, scatter_kws={"color": "blue"},␣
 ↪line_kws={"color": "red"})
plt.ylim(0)
plt.title("Correlation Between Peak RPM and Price")
plt.xlabel("Peak RPM")
plt.ylabel("Price")
plt.show()
```



PEARSON CORRELATION METHOD

[21]:
```
from scipy import stats
import pandas as pd

# Ensure 'horsepower' and 'price' are numeric
df["horsepower"] = pd.to_numeric(df["horsepower"], errors="coerce")
df["price"] = pd.to_numeric(df["price"], errors="coerce")

# Drop rows with NaN values in the relevant columns
df = df.dropna(subset=["horsepower", "price"])

# Calculate the Pearson correlation coefficient and p-value
pearson_coef, p_value = stats.pearsonr(df["horsepower"], df["price"])

# Display the results
print(f"Pearson Correlation Coefficient: {pearson_coef}")
print(f"P-value: {p_value}")
```

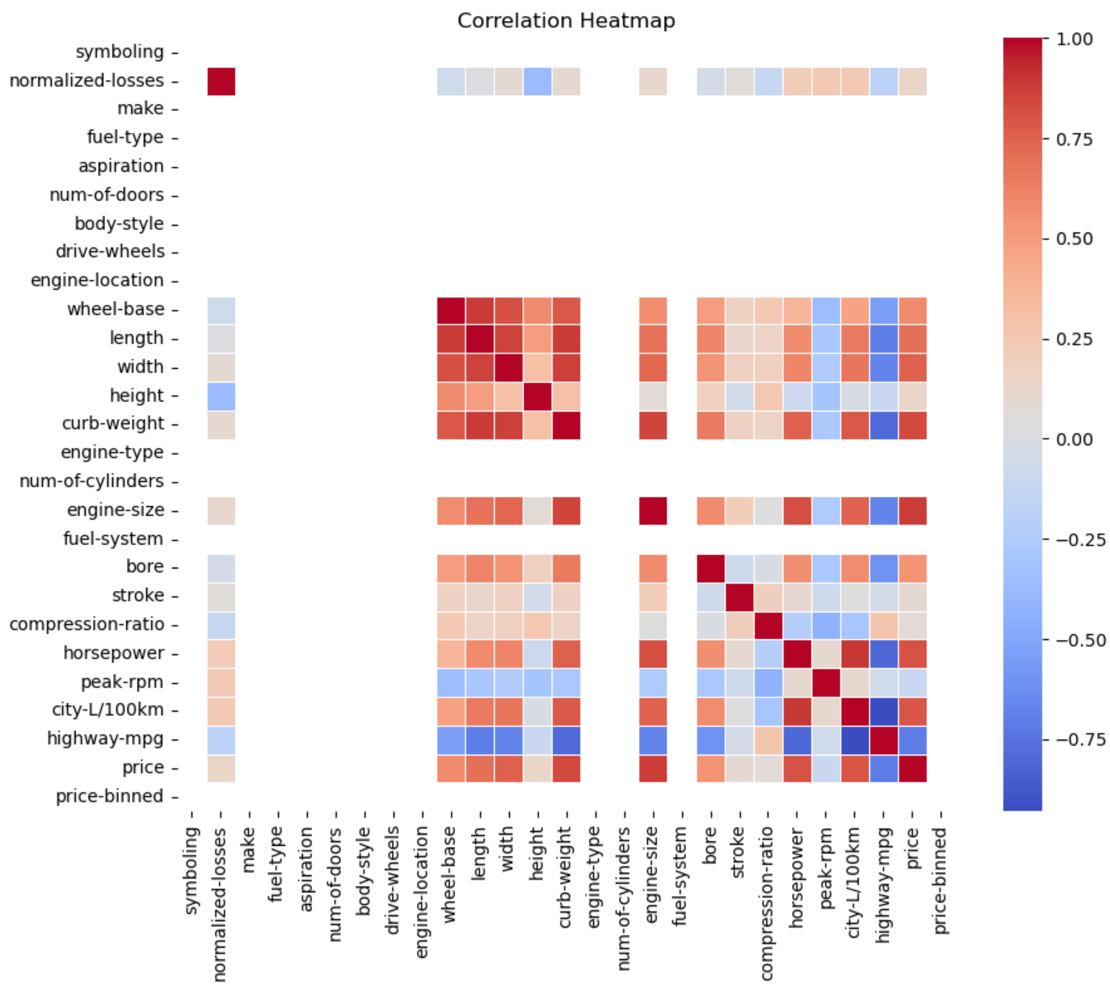
```
Pearson Correlation Coefficient: 0.8105330821322062
P-value: 1.1891278276945975e-47
```

The results indicate:

Pearson Correlation Coefficient: 0.8105 A strong positive correlation exists between horsepower and price. As horsepower increases, price tends to increase as well. P-value: 1.189×10 This extremely small p-value indicates that the correlation is statistically significant. The probability of observing this correlation by chance is effectively zero. This means horsepower is a good predictor of price in your dataset

[27]:
```python
import seaborn as sns
import matplotlib.pyplot as plt

# Convert non-numeric columns to numeric, if applicable
# Replace problematic values with NaN
df_cleaned = df.apply(pd.to_numeric, errors='coerce')

# Compute the correlation matrix (numeric columns only)
correlation_matrix = df_cleaned.corr(numeric_only=True)

# Create a heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", linewidths=0.5, fmt=".2f")

# Add title
plt.title("Correlation Heatmap")

# Display the plot
plt.show()
```

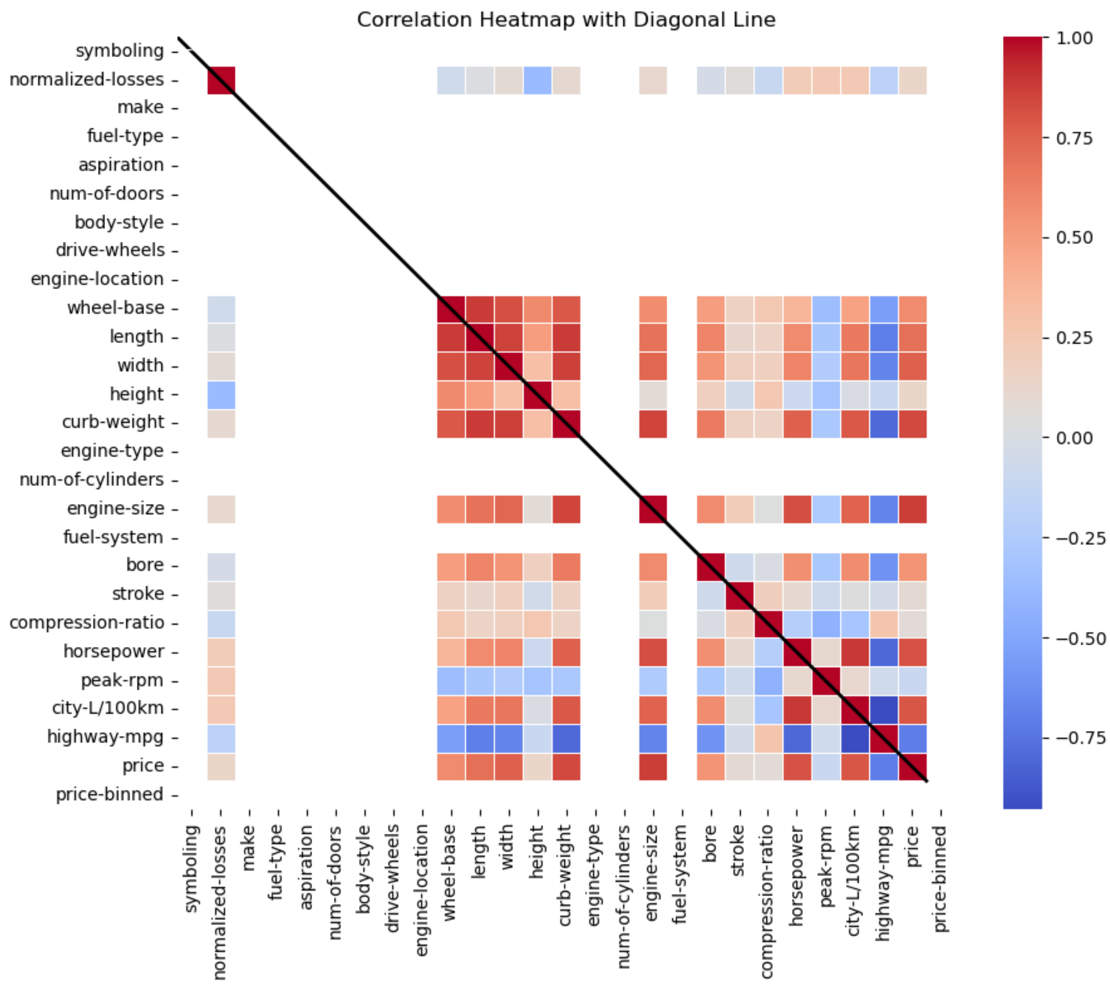
```
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/seaborn/matrix.py:260: FutureWarning: Format strings passed to
MaskedConstant are ignored, but in future may error or produce different
behavior
  annotation = ("{:" + self.fmt + "}").format(val)
```



[29]:
```python
#All of the values along the diagonal line are highly correlated
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np

# Convert non-numeric columns to numeric, if applicable
df_cleaned = df.apply(pd.to_numeric, errors='coerce')

# Compute the correlation matrix (numeric columns only)
correlation_matrix = df_cleaned.corr(numeric_only=True)

# Create a heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", linewidths=0.5,␣
↪fmt=".2f")
```

```python
# Add a diagonal line to highlight variables with high correlation (close to 1)
n = len(correlation_matrix)
plt.plot([0, n-1], [0, n-1], color='black', lw=2)

# Add title
plt.title("Correlation Heatmap with Diagonal Line")

# Display the plot
plt.show()
```

```
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/seaborn/matrix.py:260: FutureWarning: Format strings passed to
MaskedConstant are ignored, but in future may error or produce different
behavior
  annotation = ("{:" + self.fmt + "}").format(val)
```



HOW ARE THESE VARIABLES RELATED TO PRICE?

[31]:
```python
import pandas as pd

# Ensure that non-numeric columns are converted to NaN
df_cleaned = df.apply(pd.to_numeric, errors='coerce')

# Calculate the correlation matrix
correlation_matrix = df_cleaned.corr(numeric_only=True)

# Extract the correlation of all variables with "price"
price_correlation = correlation_matrix["price"]

# Display the correlation of each variable with "price"
print(price_correlation)
```

```
symboling                 NaN
normalized-losses    0.134140
make                      NaN
fuel-type                 NaN
aspiration                NaN
num-of-doors              NaN
body-style                NaN
drive-wheels              NaN
engine-location           NaN
wheel-base           0.583797
length               0.693965
width                0.753871
height               0.134990
curb-weight          0.835090
engine-type               NaN
num-of-cylinders          NaN
engine-size          0.873887
fuel-system               NaN
bore                 0.546873
stroke               0.093746
compression-ratio    0.069549
horsepower           0.810533
peak-rpm            -0.101649
city-L/100km         0.791270
highway-mpg         -0.705230
price                1.000000
price-binned              NaN
Name: price, dtype: float64
```

#Interpretation: #Positive Correlation: Variables like "engine-size", "curb-weight", "horsepower", "length", and "width" are positively correlated with "price". This suggests that as these features increase, the price also tends to increase. #Negative Correlation: Variables like "highway-mpg" and "city-mpg" are negatively correlated with "price". This means that as the miles per gallon (MPG) increases, the price tends to decrease. #Weak/No Correlation: Variables like "make" might show a low correlation, indicating a weak or no significant relationship with the price.

Convert "fuel-type" to numeric values

[33]:
```python
import pandas as pd

# Assuming 'df' is the DataFrame with your data

# Convert 'fuel-type' to numeric values (Label Encoding: "Gas" = 0, "Diesel" =
↪1)
df['fuel-type'] = df['fuel-type'].map({'gas': 0, 'diesel': 1})

# If you want to handle NaN values (optional)
df['fuel-type'] = df['fuel-type'].fillna(df['fuel-type'].mode()[0])  # Replace
↪NaN with the mode value

# Now check if the conversion worked
print(df['fuel-type'].head())
```

```
0    0
1    0
2    0
3    0
4    0
Name: fuel-type, dtype: int64
```

Calculate the correlation matrix and extract correlation with "price"

[35]:
```python
# Calculate the correlation matrix
correlation_matrix = df.corr(numeric_only=True)

# Extract the correlation of 'fuel-type' with 'price'
fuel_price_correlation = correlation_matrix['price']['fuel-type']

# Print the correlation with 'price'
print(f"Correlation between 'fuel-type' and 'price': {fuel_price_correlation}")
```

```
Correlation between 'fuel-type' and 'price': 0.10897829604216605
```

#Weak correlation (0.1 to 0.3): The fuel type does not have a strong influence on the price

[1]:
```python
import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv(r'C:\Users\dj1975\Documents\LinearRegres.csv')

# Verify the DataFrame is loaded
print(df.head())  # Show the first 5 rows to confirm the data
```

```
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
```

```
0        "3              122.0  alfa-romero       gas        std          two
1        "3              122.0  alfa-romero       gas        std          two
2        "1              122.0  alfa-romero       gas        std          two
3        "2              164.0         audi       gas        std         four
4        "2              164.0         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  fuel-system  \
0  convertible          rwd           front        88.6  …         mpfi
1  convertible          rwd           front        88.6  …         mpfi
2    hatchback          rwd           front        94.5  …         mpfi
3        sedan          fwd           front        99.8  …         mpfi
4        sedan          4wd           front        99.4  …         mpfi

   bore  stroke  compression-ratio horsepower peak-rpm  city-L/100km  \
0  3.47    2.68                9.0        111     5000     11.190476
1  3.47    2.68                9.0        111     5000     11.190476
2  2.68    3.47                9.0        154     5000     12.368421
3  3.19    3.40               10.0        102     5500      9.791667
4  3.19    3.40                8.0        115     5500     13.055556

  highway-mpg  price price-binned
0          27  13495          Low
1          27  16500          Low
2          26  16500          Low
3          30  13950          Low
4          22  17450          Low

[5 rows x 27 columns]
```

Model Development: Linear Regression, Multiple Linear Regression, Polynomial Regression and Pipelines

#Simple Linear Regression Intercept: The price when highway-mpg is 0. Coefficient: The rate of change in price for each unit change in highway-mpg.

[3]:
```
from sklearn.linear_model import LinearRegression
import numpy as np

# Create the LinearRegression object
lm = LinearRegression()

# Define the predictor (X) and target (y)
X = df[["highway-mpg"]]  # Ensure X is a 2D array
y = df["price"]  # y is a 1D array

# Fit the model
lm.fit(X, y)
```

```
# Predict using the model
Yhat = lm.predict(X)  # Pass X to the predict method

# Display results
print(f"Intercept: {lm.intercept_}")
print(f"Coefficient: {lm.coef_[0]}")
```

```
Intercept: 38423.3058581574
Coefficient: -821.7333783219254
```

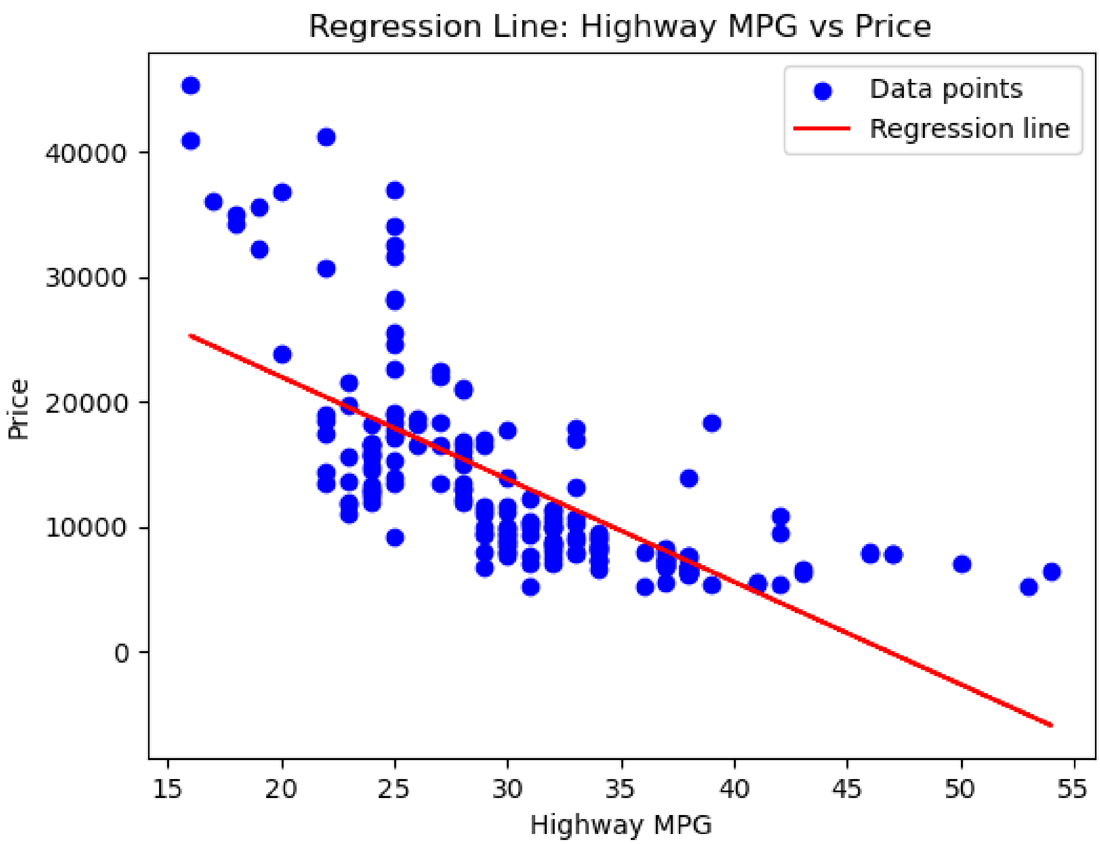
Scatter Plot:

Displays the relationship between highway-mpg and price as individual data points. Regression Line:

Uses Yhat (predicted values) for the dependent variable (price) and plots it against X (highway-mpg).

[5]:

```
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression

# Define the predictor (X) and target (y)
X = df[["highway-mpg"]]
y = df["price"]

# Create the LinearRegression object and fit the model
lm = LinearRegression()
lm.fit(X, y)

# Predict using the model
Yhat = lm.predict(X)

# Plot the data points
plt.scatter(X, y, color="blue", label="Data points")

# Plot the regression line
plt.plot(X, Yhat, color="red", label="Regression line")

# Add labels and title
plt.xlabel("Highway MPG")
plt.ylabel("Price")
plt.title("Regression Line: Highway MPG vs Price")
plt.legend()

# Show the plot
plt.show()
```



MULTIPLE LINEAR REGRESSION

[7]:
```python
from sklearn.linear_model import LinearRegression

# Initialize the model
lm = LinearRegression()

# Define predictors (Z) and target variable (y)
Z = df[["horsepower", "curb-weight", "engine-size", "highway-mpg"]]
y = df["price"]

# Fit the model
lm.fit(Z, y)

# Predict
Yhat = lm.predict(Z)

# Print coefficients and intercept for verification
print("Coefficients:", lm.coef_)
print("Intercept:", lm.intercept_)
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
/tmp/ipykernel_157/1586778756.py in ?()
      7 Z = df[["horsepower", "curb-weight", "engine-size", "highway-mpg"]]
      8 y = df["price"]
      9 
     10 # Fit the model
---> 11 lm.fit(Z, y)
     12 
     13 # Predict
     14 Yhat = lm.predict(Z)

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/base.
↪py in ?(estimator, *args, **kwargs)
   1147                 skip_parameter_validation=(
   1148                     prefer_skip_nested_validation or␣
↪global_skip_validation
   1149                 )
   1150             ):
-> 1151                 return fit_method(estimator, *args, **kwargs)

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
↪linear_model/_base.py in ?(self, X, y, sample_weight)
    674         n_jobs_ = self.n_jobs
    675 
    676         accept_sparse = False if self.positive else ["csr", "csc", "coo"]
    677 
--> 678         X, y = self._validate_data(
    679             X, y, accept_sparse=accept_sparse, y_numeric=True,␣
↪multi_output=True
    680         )
    681 

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/base.
↪py in ?(self, X, y, reset, validate_separately, cast_to_ndarray,␣
↪**check_params)
    617                 if "estimator" not in check_y_params:
    618                     check_y_params = {**default_check_params,␣
↪**check_y_params}
    619                 y = check_array(y, input_name="y", **check_y_params)
    620             else:
--> 621                 X, y = check_X_y(X, y, **check_params)
    622             out = X, y
    623 
    624         if not no_val_X and check_params.get("ensure_2d", True):
```

```
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
↪utils/validation.py in ?(X, y, accept_sparse, accept_large_sparse, dtype,␣
↪order, copy, force_all_finite, ensure_2d, allow_nd, multi_output,␣
↪ensure_min_samples, ensure_min_features, y_numeric, estimator)
   1143         raise ValueError(
   1144             f"{estimator_name} requires y to be passed, but the target ␣
↪is None"
   1145         )
   1146
-> 1147     X = check_array(

   1148         X,
   1149         accept_sparse=accept_sparse,
   1150         accept_large_sparse=accept_large_sparse,

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
↪utils/validation.py in ?(array, accept_sparse, accept_large_sparse, dtype,␣
↪order, copy, force_all_finite, ensure_2d, allow_nd, ensure_min_samples,␣
↪ensure_min_features, estimator, input_name)
    914                 )
    915                 array = xp.astype(array, dtype, copy=False)
    916             else:
    917                 array = _asarray_with_order(array, order=order,␣
↪dtype=dtype, xp=xp)
--> 918         except ComplexWarning as complex_warning:
    919             raise ValueError(
    920                 "Complex data not supported\n{}\n".format(array)
    921             ) from complex_warning

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
↪utils/_array_api.py in ?(array, dtype, order, copy, xp)
    376         # Use NumPy API to support order
    377         if copy is True:
    378             array = numpy.array(array, order=order, dtype=dtype)
    379         else:
--> 380             array = numpy.asarray(array, order=order, dtype=dtype)
    381
    382         # At this point array is a NumPy ndarray. We convert it to an␣
↪array
    383         # container that is consistent with the input's namespace.

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/pandas/core/
↪generic.py in ?(self, dtype)
   2082     def __array__(self, dtype: npt.DTypeLike | None = None) -> np.
↪ndarray:
   2083         values = self._values
-> 2084         arr = np.asarray(values, dtype=dtype)
   2085         if (
   2086             astype_is_view(values.dtype, arr.dtype)
```

```
   2087                 and using_copy_on_write()

ValueError: could not convert string to float: '?'
```

The error indicates that your dataset contains non-numeric values ('?') in one or more of the predictor columns (Z) or the target column (price). Linear regression models require numeric data.

STEPS TO RESOLVE

[9]:
```
# Replace '?' with NaN in the DataFrame
df.replace('?', np.nan, inplace=True)
```

[11]:
```
#Convert Columns to Numeric
df["horsepower"] = pd.to_numeric(df["horsepower"], errors='coerce')
df["curb-weight"] = pd.to_numeric(df["curb-weight"], errors='coerce')
df["engine-size"] = pd.to_numeric(df["engine-size"], errors='coerce')
df["highway-mpg"] = pd.to_numeric(df["highway-mpg"], errors='coerce')
df["price"] = pd.to_numeric(df["price"], errors='coerce')
```

[13]:
```
# Drop rows with missing values
df.dropna(subset=["horsepower", "curb-weight", "engine-size", "highway-mpg",␣
 ↪"price"], inplace=True)
```

[15]:
```
#Run the Model Again
from sklearn.linear_model import LinearRegression

lm = LinearRegression()

# Define predictors and target variable
Z = df[["horsepower", "curb-weight", "engine-size", "highway-mpg"]]
y = df["price"]

# Fit the model
lm.fit(Z, y)

# Predict
Yhat = lm.predict(Z)

# Print results
print("Coefficients:", lm.coef_)
print("Intercept:", lm.intercept_)
```

```
Coefficients: [53.27878556  4.66217408 82.22948394 35.5175845 ]
Intercept: -15700.573979039307
```

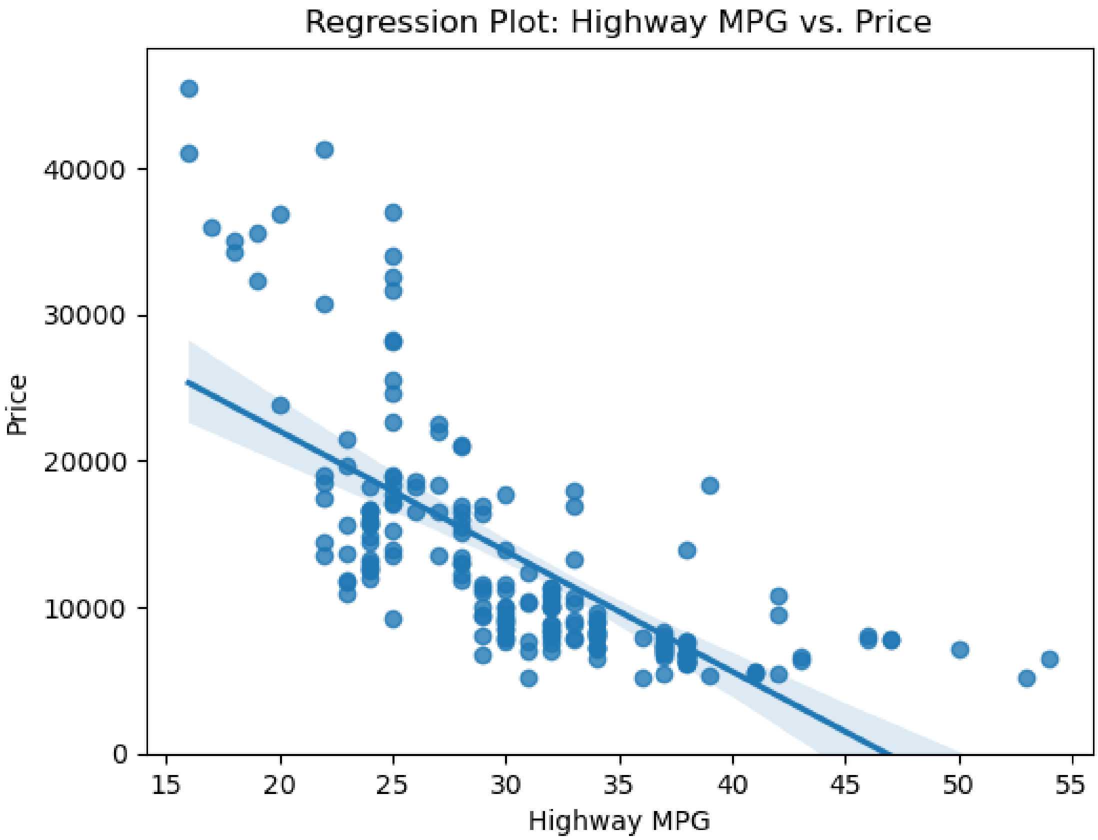
[17]:
```
# Coefficients and Intercept
coefficients = lm.coef_
intercept = lm.intercept_
```

```
# Predictor names
predictor_names = ["horsepower", "curb-weight", "engine-size", "highway-mpg"]

# Display interpretations
print(f"Intercept: The predicted price when all predictors are 0 is ${intercept:
 ↪.2f}.")

for i, coef in enumerate(coefficients):
    if coef > 0:
        direction = "increase"
    else:
        direction = "decrease"
    print(f"Coefficient for {predictor_names[i]}: "
          f"For every additional unit of {predictor_names[i]}, "
          f"the price is expected to {direction} by ${abs(coef):.2f}, holding␣
 ↪other variables constant.")
```

```
Intercept: The predicted price when all predictors are 0 is $-15700.57.
Coefficient for horsepower: For every additional unit of horsepower, the price
is expected to increase by $53.28, holding other variables constant.
Coefficient for curb-weight: For every additional unit of curb-weight, the price
is expected to increase by $4.66, holding other variables constant.
Coefficient for engine-size: For every additional unit of engine-size, the price
is expected to increase by $82.23, holding other variables constant.
Coefficient for highway-mpg: For every additional unit of highway-mpg, the price
is expected to increase by $35.52, holding other variables constant.
```

REGRESSION PLOT

[19]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Create the regression plot
sns.regplot(x="highway-mpg", y="price", data=df)

# Set the y-axis limit
plt.ylim(0)

# Add labels and title
plt.xlabel("Highway MPG")
plt.ylabel("Price")
plt.title("Regression Plot: Highway MPG vs. Price")

# Display the plot
plt.show()
```



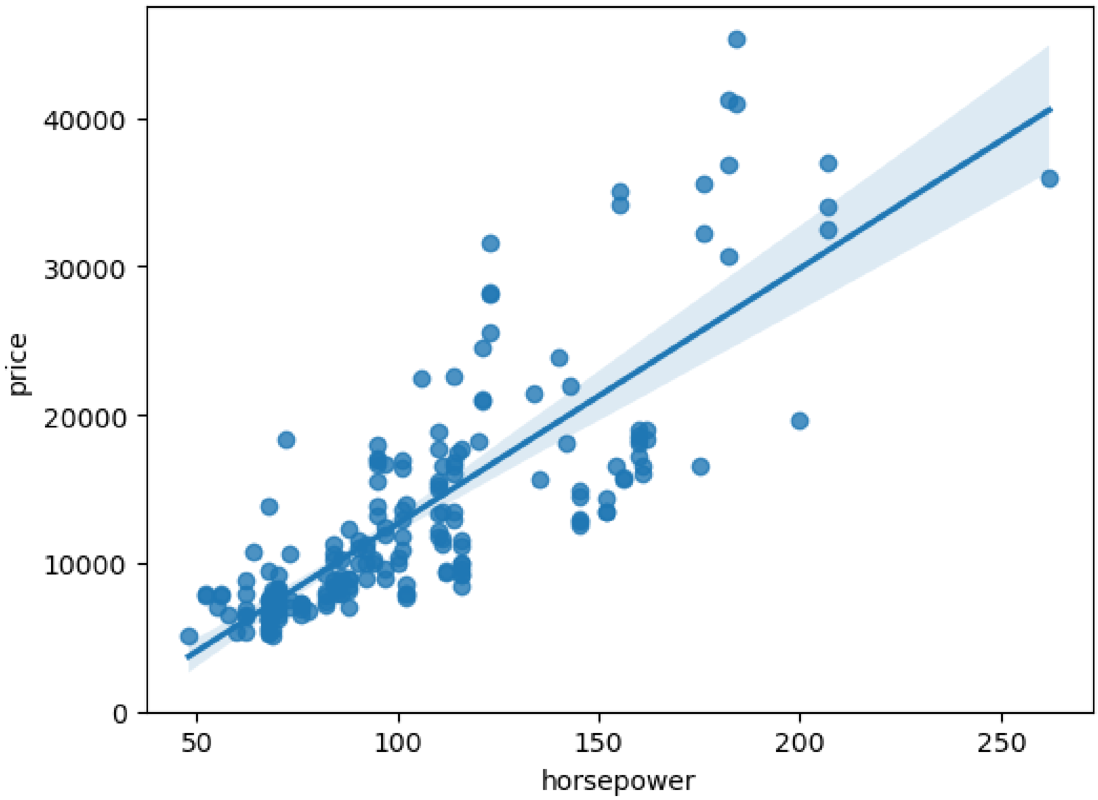
If the dots are scattered all over the place with no clear pattern, this suggests that the relationship between highway-mpg and price is weak or not linear. In other words:

Weak correlation: There is no strong linear relationship between highway-mpg and price, meaning highway-mpg does not strongly predict car prices. Non-linear relationship: The relationship may be more complex, requiring other factors to better explain the price variation. Outliers or noise: The data may contain noise or outliers that obscure any potential pattern. This insight means that highway-mpg alone is not a good predictor of price.

#Regression Plot: X="horsepower", y="price"

[21]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Create the regression plot
sns.regplot(x="horsepower", y="price", data=df)

# Set the y-axis limit to start from 0
plt.ylim(0)

# Display the plot
```

```
plt.show()
```

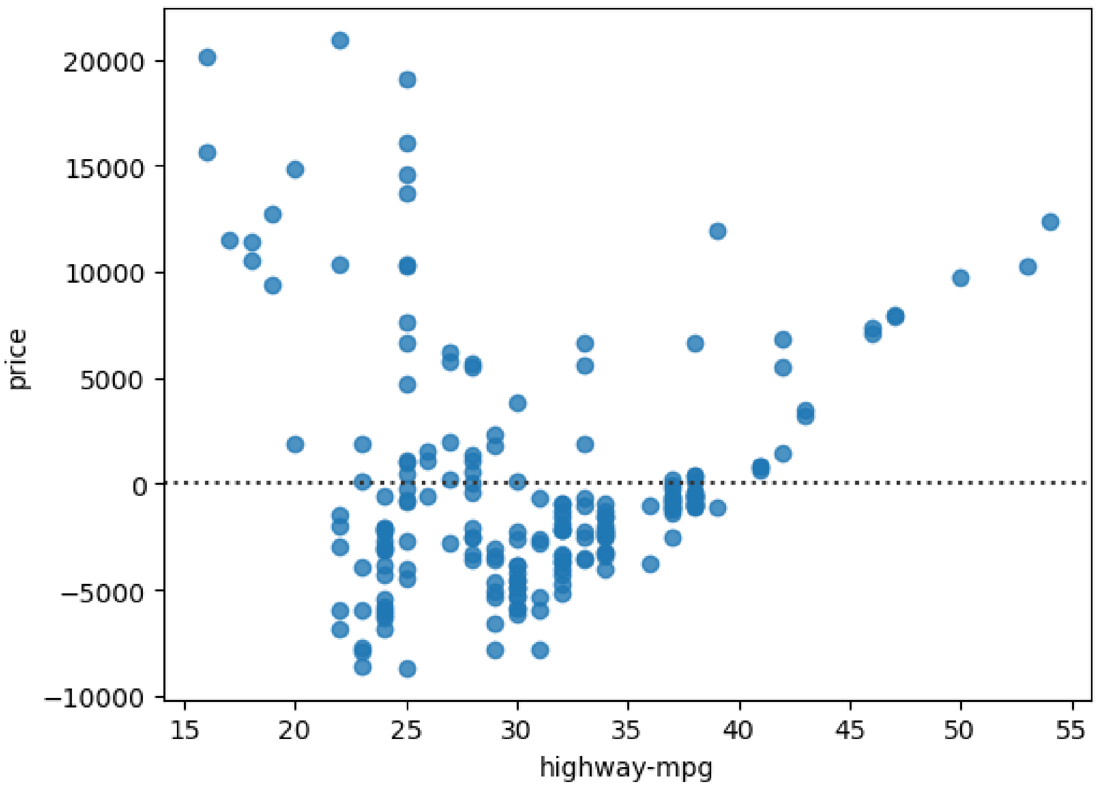


While the regression plot provides insight into the relationship between horsepower and price, its reliability depends on factors such as data distribution, outliers, and the underlying assumptions of linear regression. If the data is mostly clustered around a certain range (e.g., 0-125 horsepower), the model may not fully capture the relationship for higher values.

RESIDUAL PLOT

This plot will display the residuals (the difference between actual and predicted values) against the "highway-mpg" variable, allowing you to assess the fit of the linear regression model and check for patterns that might suggest non-linearity or heteroscedasticity.

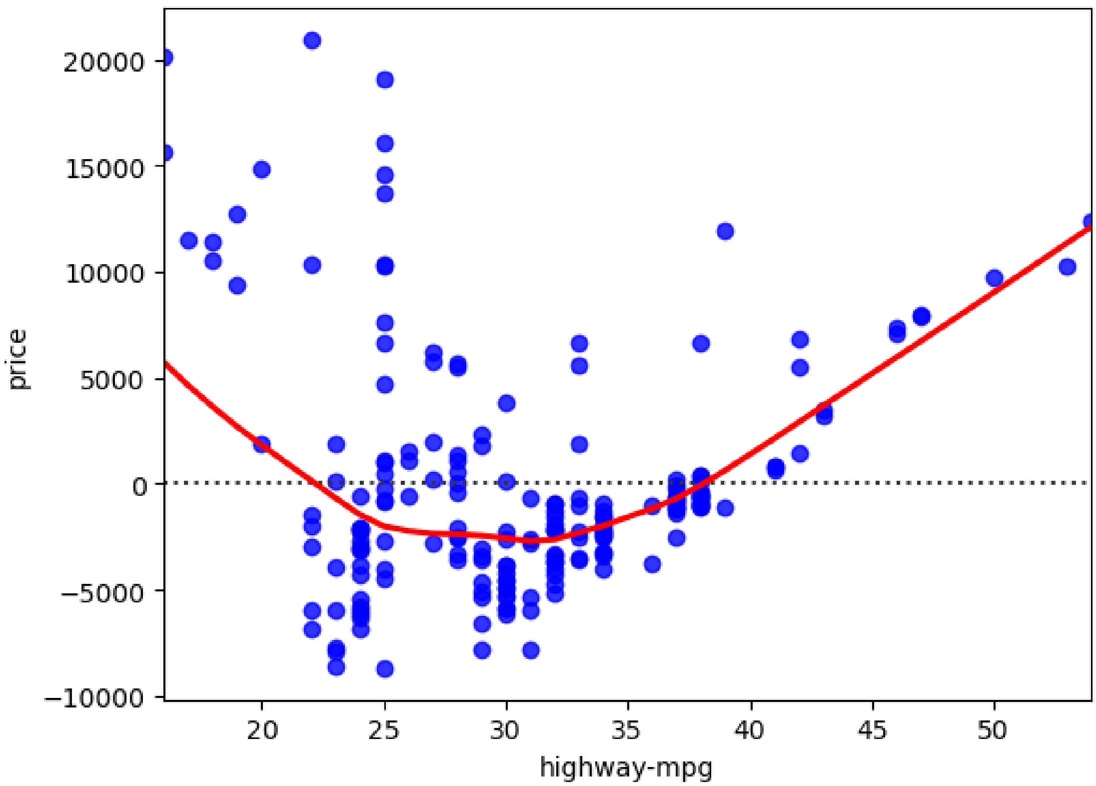
[25]:
```
import seaborn as sns
import matplotlib.pyplot as plt

sns.residplot(x="highway-mpg", y="price", data=df)
plt.show()
```



When the dots are not centered around 0, it might suggest:

Model Misspecification: The model might not be the best fit for the data, or the features selected may not fully capture the relationship with the target variable. Non-linearity: If the relationship between the predictors and the target is non-linear, a linear regression model might not capture it accurately. Outliers or influential points: Some data points might disproportionately affect the model's predictions, leading to residuals that aren't centered around 0.

[27]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Create a residual plot
sns.residplot(x=df["highway-mpg"], y=df["price"], color="blue", lowess=True,␣
 ↪line_kws={'color': 'red'})

# Show plot
plt.show()
```



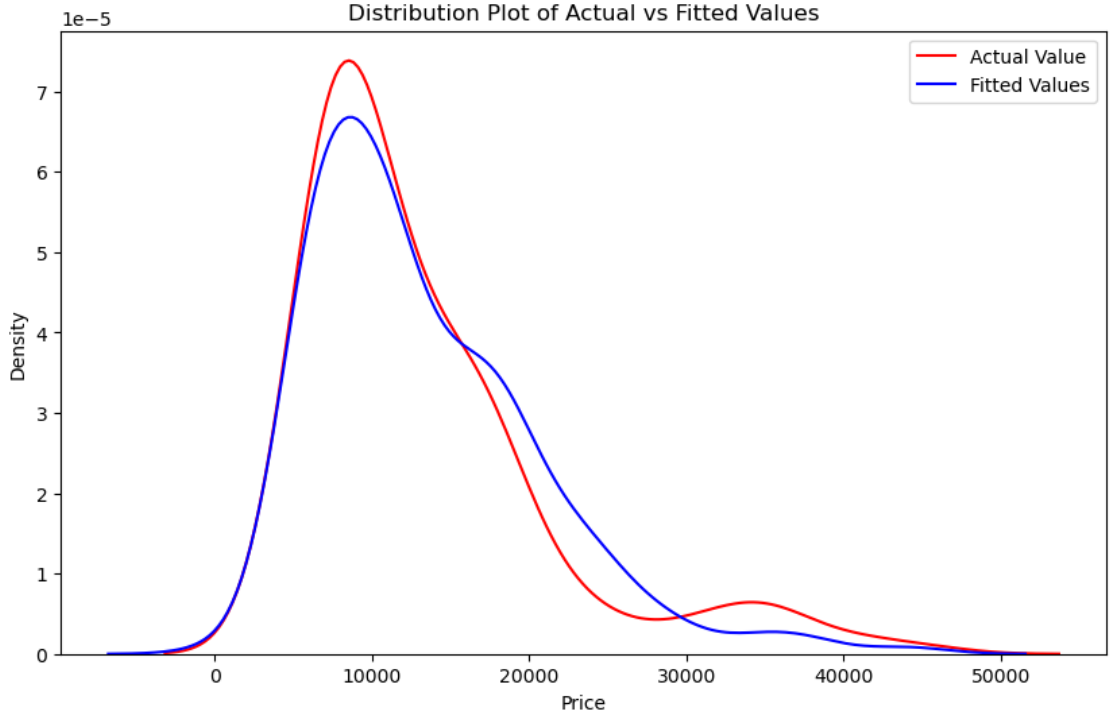
The curved red line represents a lowess (locally weighted scatterplot smoothing) line. It is used to show any potential non-linear relationship between the independent variable (highway-mpg) and the residuals. If the line is curved, it suggests that the relationship might not be purely linear, indicating that a linear model may not be the best fit for the data.

DISTRIBUTION PLOT

[29]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Create the distribution plot for price and predicted values
plt.figure(figsize=(10, 6))
ax1 = sns.distplot(df["price"], hist=False, color="r", label="Actual Value")
sns.distplot(Yhat, hist=False, color="b", label="Fitted Values", ax=ax1)

# Add labels and title
plt.xlabel('Price')
plt.ylabel('Density')
plt.title('Distribution Plot of Actual vs Fitted Values')
plt.legend()
plt.show()
```

```
/tmp/ipykernel_157/182394368.py:6: 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
similar flexibility) or `kdeplot` (an axes-level function for kernel density
plots).

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

  ax1 = sns.distplot(df["price"], hist=False, color="r", label="Actual Value")
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/seaborn/_oldcore.py:1119: FutureWarning: use_inf_as_na option is
deprecated and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
/tmp/ipykernel_157/182394368.py:7: 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
similar flexibility) or `kdeplot` (an axes-level function for kernel density
plots).

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

  sns.distplot(Yhat, hist=False, color="b", label="Fitted Values", ax=ax1)
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/seaborn/_oldcore.py:1119: FutureWarning: use_inf_as_na option is
deprecated and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



The red line represents the distribution of the actual values of the price. The blue line represents the distribution of the predicted (fitted) values from the regression model. This plot helps to visually assess how well the model's predicted values align with the actual values. If the two distributions are similar, it suggests that the model is doing a good job at capturing the underlying data.

Fitted values (also known as predicted values) are the values that your regression model predicts based on the input features (e.g., horsepower, curb-weight, engine-size, etc.). These are the outputs generated by the model when it tries to estimate the price of the car using the regression equation.

Key difference: Actual values: What the data actually shows (the observed prices in your dataset). Fitted values: What the model predicts for those same data points based on the patterns it learned from the data.

Purpose of comparing them: By comparing the actual and fitted values, you can evaluate how well the model is performing. If the fitted values closely match the actual values, it indicates that the model has captured the underlying relationship between the features and the target variable well.

Polynomial Regression and Pipelines: Used when a linear model is not the best fit for your data

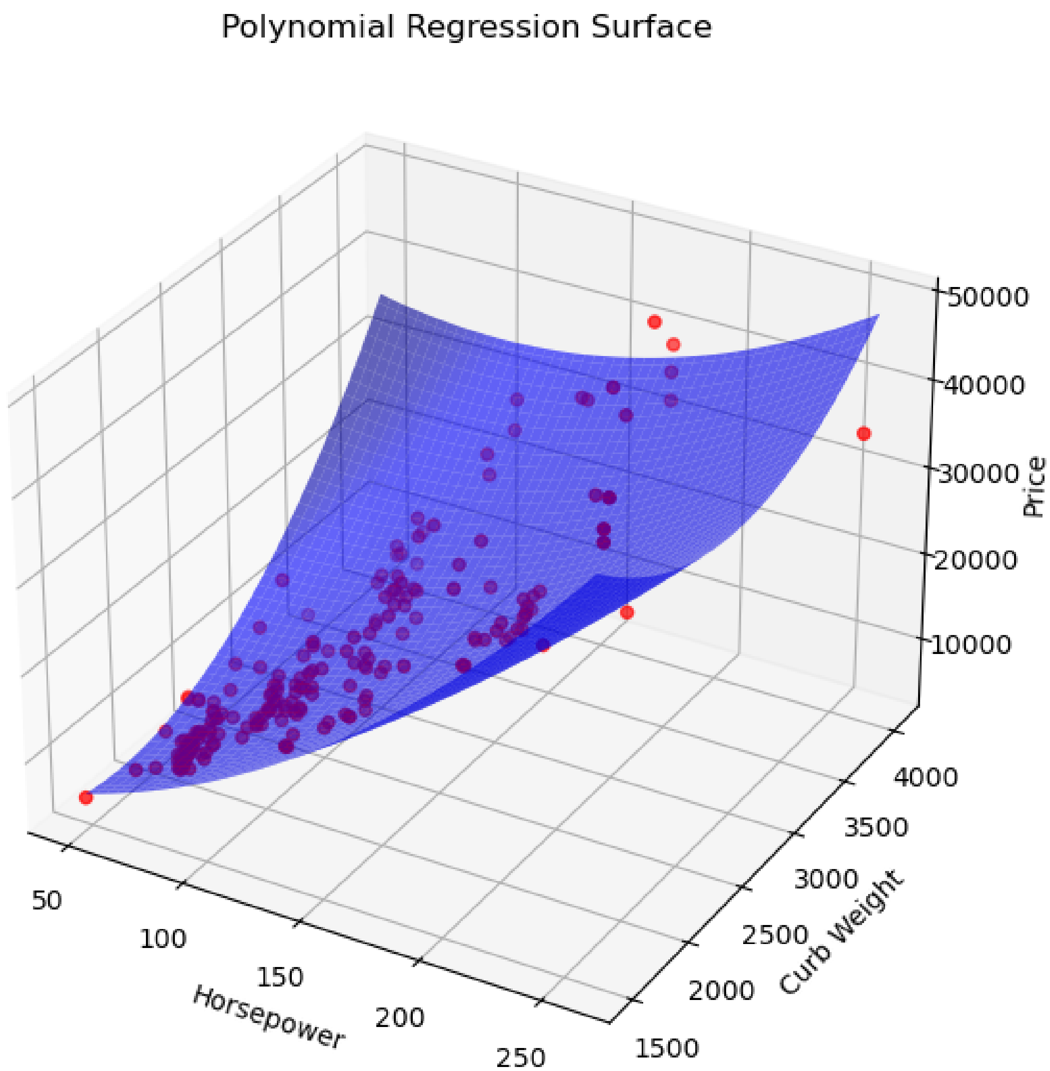
[31]:
```
from sklearn.preprocessing import PolynomialFeatures
pr = PolynomialFeatures(degree=2, include_bias=False)

# Transform the input features (e.g., horsepower, curb-weight) into polynomial␣
↪features
x_poly = pr.fit_transform(df[["horsepower", "curb-weight"]])
```

```
# You can then use these features for regression
```

Visualize the Polynomial Regression Model

[33]:
```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression

# Fit the polynomial regression model
pr_model = LinearRegression()
pr_model.fit(x_poly, df["price"])

# Create a meshgrid for plotting the surface
x_range = np.linspace(df["horsepower"].min(), df["horsepower"].max(), 100)
y_range = np.linspace(df["curb-weight"].min(), df["curb-weight"].max(), 100)
X, Y = np.meshgrid(x_range, y_range)
Z = pr_model.predict(pr.transform(np.c_[X.ravel(), Y.ravel()])).reshape(X.shape)

# Plot the actual data points and the polynomial regression surface
fig = plt.figure(figsize=(10, 7))
ax = fig.add_subplot(111, projection='3d')

# Scatter plot of the actual data points
ax.scatter(df["horsepower"], df["curb-weight"], df["price"], color='r',␣
↪label='Actual Data')

# Plot the polynomial regression surface
ax.plot_surface(X, Y, Z, color='b', alpha=0.6, label='Polynomial Fit')

ax.set_xlabel('Horsepower')
ax.set_ylabel('Curb Weight')
ax.set_zlabel('Price')
ax.set_title('Polynomial Regression Surface')
plt.show()
```

```
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py:464: UserWarning: X does not have valid feature names,
but PolynomialFeatures was fitted with feature names
  warnings.warn(
```



Recommendations: Normalize or Scale Features: Consider scaling curb weight and other numerical features so that they are on similar scales. This can improve the model's performance, especially when dealing with polynomial regression or other machine learning models.

Segment the Data: If you're trying to predict the price for a broader range of vehicles, consider breaking the data into subsets (e.g., by vehicle type) or increasing the diversity of your data, particularly including cars with higher curb weights or higher prices.

Interaction Terms: Add interaction terms between curb weight and other features (such as horsepower) to explore non-linear relationships. For example, higher curb weight might interact with horsepower to influence the price.

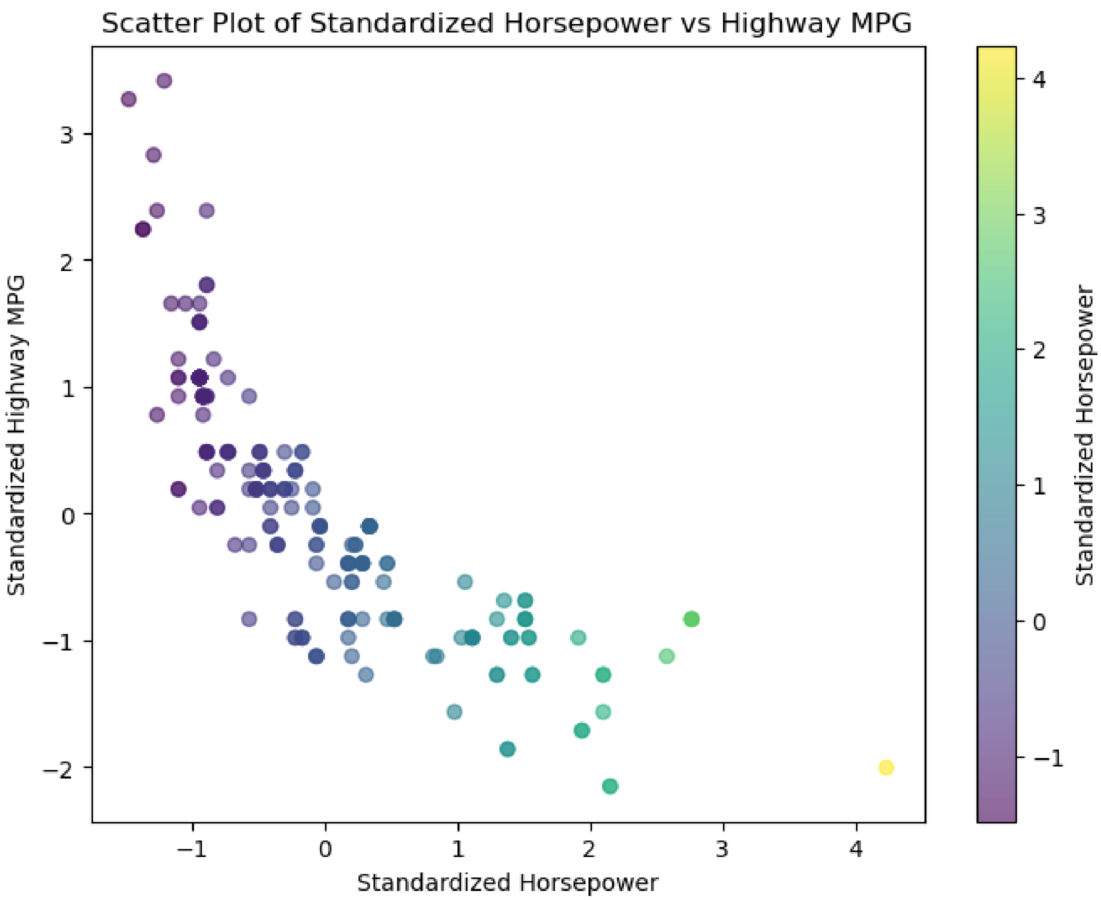
Polynomial Features: If you are using polynomial regression, you can experiment with increasing the degree of polynomial features to capture the more subtle non-linear effects between curb weight and price.

PRE-PROCESSING

[37]:
```
from sklearn.preprocessing import StandardScaler

# Initialize the StandardScaler
scaler = StandardScaler()

# Fit and transform the data using the correct DataFrame name (df)
x_scaled = scaler.fit_transform(df[["horsepower", "highway-mpg"]])

# Now x_scaled contains the standardized values
```

Visualize the relationship between horsepower and highway mpg

[41]:
```
import matplotlib.pyplot as plt
import numpy as np

# Create a scatter plot with a color map based on horsepower values
plt.figure(figsize=(8, 6))
scatter = plt.scatter(x_scaled[:, 0], x_scaled[:, 1], c=x_scaled[:, 0],␣
↪cmap='viridis', alpha=0.6)
plt.xlabel("Standardized Horsepower")
plt.ylabel("Standardized Highway MPG")
plt.title("Scatter Plot of Standardized Horsepower vs Highway MPG")
plt.colorbar(scatter, label="Standardized Horsepower")  # Add a color bar
plt.show()
```



Check the correlation between horsepower and highway mpg

[43]:
```
correlation = df["horsepower"].corr(df["highway-mpg"])
print(f"Correlation between horsepower and highway-mpg: {correlation}")
```

```
Correlation between horsepower and highway-mpg: -0.8045960570124867
```

The correlation of -0.80 between horsepower and highway-mpg indicates a strong negative correlation. This means that, generally:

As horsepower increases, highway-mpg tends to decrease. Cars with higher horsepower typically have lower fuel efficiency (in terms of highway miles per gallon).

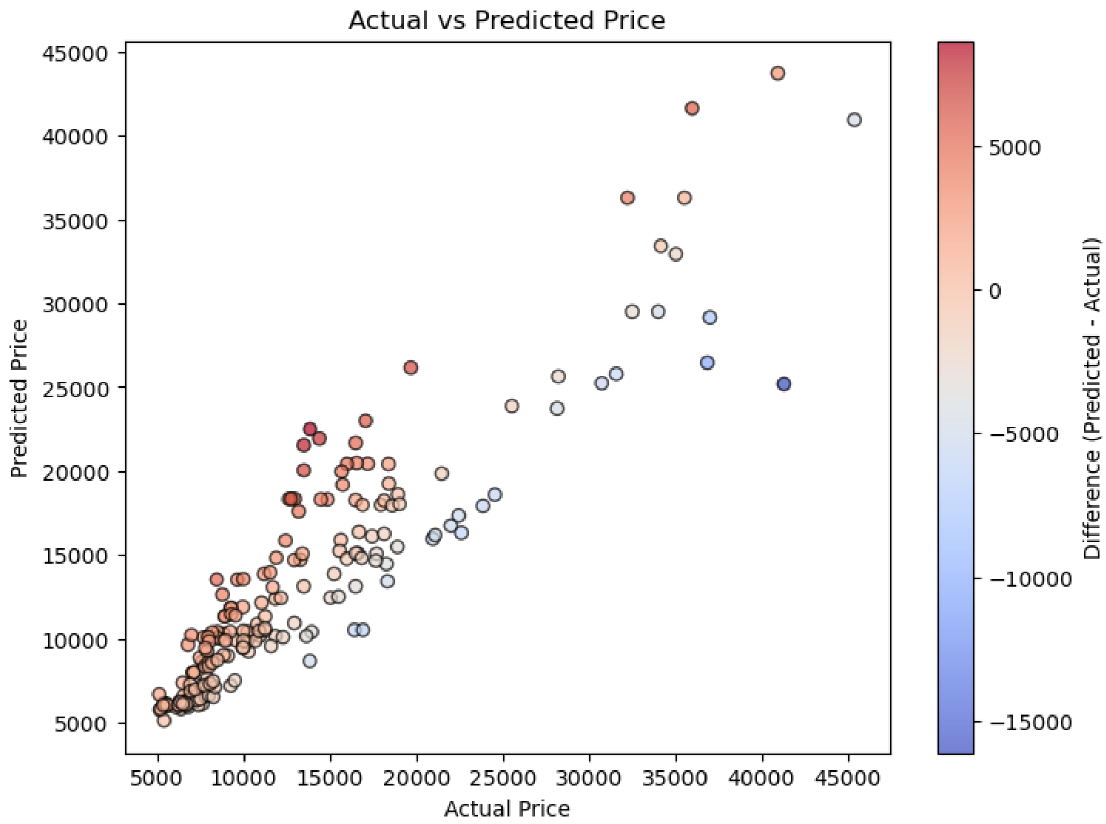
PIPELINES

[45]:
```
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import Pipeline

# Create the pipeline
```

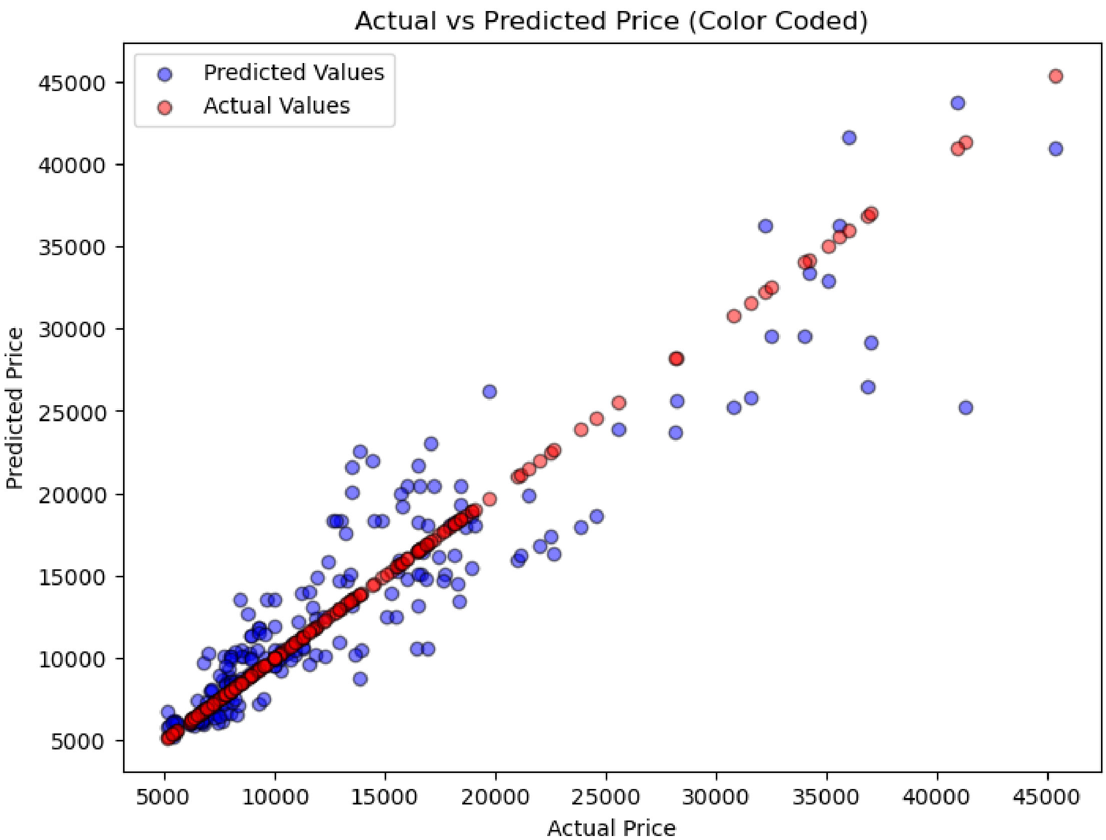
```python
pipe = Pipeline([
    ('polynomial', PolynomialFeatures(degree=2)),
    ('scale', StandardScaler()),
    ('model', LinearRegression())
])

# Fit the pipeline to the data
pipe.fit(df[["horsepower", "curb-weight", "engine-size", "highway-mpg"]],␣
 ↪df["price"])

# Predict using the pipeline
Yhat = pipe.predict(df[["horsepower", "curb-weight", "engine-size",␣
 ↪"highway-mpg"]])
```

VISUALIZE THE PIPELINE

[49]:
```python
import matplotlib.pyplot as plt
import numpy as np

# Scatter plot of actual vs predicted values
plt.figure(figsize=(8,6))
plt.scatter(df['price'], Yhat, c=Yhat - df['price'], cmap='coolwarm',␣
 ↪edgecolor='k', alpha=0.7)
plt.colorbar(label='Difference (Predicted - Actual)')  # Color bar to show the␣
 ↪difference scale
plt.xlabel('Actual Price')
plt.ylabel('Predicted Price')
plt.title('Actual vs Predicted Price')
plt.show()
```



[59]:
```
# Actual vs Predicted Values with Color
plt.figure(figsize=(8,6))

# Plot actual values
plt.scatter(df['price'], Yhat, c='blue', label='Predicted Values',␣
  ↪edgecolor='k', alpha=0.5)

# Plot predicted values
plt.scatter(df['price'], df['price'], c='red', label='Actual Values',␣
  ↪edgecolor='k', alpha=0.5)

plt.xlabel('Actual Price')
plt.ylabel('Predicted Price')
plt.title('Actual vs Predicted Price (Color Coded)')
plt.legend()
plt.show()
```



Based on the observation that the blue dots (predicted values) are concentrated near the bottom left of the graph, the model is likely predicting lower-priced cars more accurately. This suggests that the model is biased toward predicting lower values and may not be capturing the patterns that would allow it to predict higher-priced cars effectively. The under-prediction for higher-priced vehicles indicates that the model performs better for the lower price range and struggles with the higher price range.

To improve predictions for higher-priced cars, you may need to:

Enhance the model: Consider using more complex models (e.g., polynomial regression, decision trees) that can capture non-linear relationships. Add more relevant features: Incorporate other factors that might influence the price of a car, which could help the model generalize better. Handle outliers: Investigate and address outliers that may be affecting the model's ability to predict accurately for higher prices.

TEST MODEL PERFROMANCE

[67]:
```
from sklearn.tree import DecisionTreeRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt
```

```
# Features and target variable
X = df[["horsepower", "curb-weight", "engine-size", "highway-mpg"]]  # Input␣
↪features
y = df["price"]  # Target variable

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=42)

# Initialize and train the Decision Tree model
dt_model = DecisionTreeRegressor(random_state=42)
dt_model.fit(X_train, y_train)

# Make predictions
y_pred = dt_model.predict(X_test)

# Evaluate the model performance
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")

# Visualizing the Decision Tree (Optional)
from sklearn.tree import plot_tree
plt.figure(figsize=(12, 8))
plot_tree(dt_model, filled=True, feature_names=X.columns, fontsize=10)
plt.show()
```

```
Mean Squared Error: 9542543.475
```

```
---------------------------------------------------------------------------
InvalidParameterError                     Traceback (most recent call last)
Cell In[67], line 27
     25 from sklearn.tree import plot_tree
     26 plt.figure(figsize=(12, 8))
---> 27 plot_tree(dt_model, filled=True, feature_names=X.columns, fontsize=10)
     28 plt.show()

File /opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
↪utils/_param_validation.py:201, in validate_params.<locals>.decorator.<locals>.
↪wrapper(*args, **kwargs)
    198 to_ignore += ["self", "cls"]
    199 params = {k: v for k, v in params.arguments.items() if k not in␣
↪to_ignore}
--> 201 validate_parameter_constraints(
    202     parameter_constraints, params, caller_name=func.__qualname__
    203 )
    205 try:
    206     with config_context(
```

```
    207         skip_parameter_validation=(
    208             prefer_skip_nested_validation or global_skip_validation
    209         )
    210     ):

File /opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
 ↪utils/_param_validation.py:95, in␣
 ↪validate_parameter_constraints(parameter_constraints, params, caller_name)
     89 else:
     90     constraints_str = (
     91         f"{', '.join([str(c) for c in constraints[:-1]])} or"
     92         f" {constraints[-1]}"
     93     )
---> 95 raise InvalidParameterError(
     96     f"The {param_name!r} parameter of {caller_name} must be"
     97     f" {constraints_str}. Got {param_val!r} instead."
     98 )

InvalidParameterError: The 'feature_names' parameter of plot_tree must be an␣
 ↪instance of 'list' or None. Got Index(['horsepower', 'curb-weight',␣
 ↪'engine-size', 'highway-mpg'], dtype='object') instead.
```

```
<Figure size 1200x800 with 0 Axes>
```

A Mean Squared Error (MSE) of 9542543.475 indicates that, on average, the difference between the predicted prices and the actual prices is substantial. This value suggests that the model is not performing very well, and there might be some room for improvement. A lower MSE would indicate better accuracy of the model in predicting the target variable (price).

Possible actions to improve: Hyperparameter Tuning: You could try tuning the decision tree's hyperparameters (like max_depth, min_samples_split, min_samples_leaf) to prevent overfitting or underfitting.

Feature Engineering: Add or transform features to better capture the patterns in the data.

Different Model: Consider trying other models like Random Forest, Gradient Boosting, or even Linear Regression, and compare their performance.

Cross-Validation: Use cross-validation to get a more robust estimate of model performance and avoid overfitting to the train-test split.

TEST FOR OVERFITTING

[69]:

```
from sklearn.metrics import mean_squared_error

# Predict on training data
train_predictions = dt_model.predict(X_train)
train_mse = mean_squared_error(y_train, train_predictions)

# Predict on testing data
test_predictions = dt_model.predict(X_test)
```

```
test_mse = mean_squared_error(y_test, test_predictions)

print(f"Training MSE: {train_mse}")
print(f"Testing MSE: {test_mse}")
```

```
Training MSE: 59801.29035639413
Testing MSE: 9542543.475
```

The large difference between the training MSE (59,801) and the testing MSE (9,542,543) indicates overfitting. Specifically:

Training MSE is very low, which means the model is performing well on the training data. Testing MSE is much higher, which suggests that the model is not generalizing well to new, unseen data and is instead memorizing the training data. In this case, the decision tree is likely overfitting, meaning it is too complex and too closely fitted to the training data, resulting in poor performance on the testing data.

Solutions to address overfitting: 1)Pruning the tree: Limit the depth of the decision tree by setting a maximum depth (max_depth) or a minimum number of samples per leaf (min_samples_leaf). 2)Cross-validation: Use cross-validation to ensure the model generalizes well. 3)Increase training data: More data can help the model generalize better. 4)Regularization: Apply regularization techniques to reduce the complexity of the model.

INCREASE TEST_SIZE FROM 0.2 to 0.3

[71]:
```
from sklearn.tree import DecisionTreeRegressor
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt

# Features and target variable
X = df[["horsepower", "curb-weight", "engine-size", "highway-mpg"]]  # Input␣
↪features
y = df["price"]  # Target variable

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,␣
↪random_state=42)

# Initialize and train the Decision Tree model
dt_model = DecisionTreeRegressor(random_state=42)
dt_model.fit(X_train, y_train)

# Make predictions
y_pred = dt_model.predict(X_test)

# Evaluate the model performance
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")
```

```python
# Visualizing the Decision Tree (Optional)
from sklearn.tree import plot_tree
plt.figure(figsize=(12, 8))
plot_tree(dt_model, filled=True, feature_names=X.columns, fontsize=10)
plt.show()
```

```
Mean Squared Error: 8292717.5
```

```
---------------------------------------------------------------------------
InvalidParameterError                     Traceback (most recent call last)
Cell In[71], line 27
     25 from sklearn.tree import plot_tree
     26 plt.figure(figsize=(12, 8))
---> 27 plot_tree(dt_model, filled=True, feature_names=X.columns, fontsize=10)
     28 plt.show()

File /opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
↪utils/_param_validation.py:201, in validate_params.<locals>.decorator.<locals>.
↪wrapper(*args, **kwargs)
    198 to_ignore += ["self", "cls"]
    199 params = {k: v for k, v in params.arguments.items() if k not in␣
↪to_ignore}
--> 201 validate_parameter_constraints(
    202     parameter_constraints, params, caller_name=func.__qualname__
    203 )
    205 try:
    206     with config_context(
    207         skip_parameter_validation=(
    208             prefer_skip_nested_validation or global_skip_validation
    209         )
    210     ):

File /opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
↪utils/_param_validation.py:95, in␣
↪validate_parameter_constraints(parameter_constraints, params, caller_name)
     89 else:
     90     constraints_str = (
     91         f"{', '.join([str(c) for c in constraints[:-1]])} or"
     92         f" {constraints[-1]}"
     93     )
---> 95 raise InvalidParameterError(
     96     f"The {param_name!r} parameter of {caller_name} must be"
     97     f" {constraints_str}. Got {param_val!r} instead."
     98 )
```

```
InvalidParameterError: The 'feature_names' parameter of plot_tree must be an␣
↪instance of 'list' or None. Got Index(['horsepower', 'curb-weight',␣
↪'engine-size', 'highway-mpg'], dtype='object') instead.
```

```
<Figure size 1200x800 with 0 Axes>
```

#Larger Test Size Impact: With a larger test set (0.3), you are getting a more robust measure of the model's performance. The fact that the test MSE decreases when you increase the test size suggests that the model may not be overfitting as severely as it might have with a smaller test set. Conclusion:

#The change in test size affects the variance in your performance metric. While both MSE values are still high, it does not seem like overfitting is the primary issue here, especially since the MSE is more consistent with a larger test set (0.3). However, it is still possible the model is not generalizing well, and you may want to explore further adjustments like pruning or hyperparameter tuning to improve the model's performance.

[1]:
```
import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv(r'C:\Users\dj1975\Documents\LinearRegres.csv')

# Verify the DataFrame is loaded
print(df.head())  # Show the first 5 rows to confirm the data
```

```
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3              122.0  alfa-romero       gas        std          two
1        "3              122.0  alfa-romero       gas        std          two
2        "1              122.0  alfa-romero       gas        std          two
3        "2              164.0         audi       gas        std         four
4        "2              164.0         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  fuel-system  \
0  convertible          rwd           front        88.6  …         mpfi
1  convertible          rwd           front        88.6  …         mpfi
2    hatchback          rwd           front        94.5  …         mpfi
3        sedan          fwd           front        99.8  …         mpfi
4        sedan          4wd           front        99.4  …         mpfi

   bore  stroke  compression-ratio horsepower peak-rpm  city-L/100km  \
0  3.47    2.68                9.0        111     5000     11.190476
1  3.47    2.68                9.0        111     5000     11.190476
2  2.68    3.47                9.0        154     5000     12.368421
3  3.19    3.40               10.0        102     5500      9.791667
4  3.19    3.40                8.0        115     5500     13.055556

  highway-mpg  price price-binned
0          27  13495          Low
1          27  16500          Low
```

```
2           26  16500          Low
3           30  13950          Low
4           22  17450          Low

[5 rows x 27 columns]
```

MEASURES FOR IN-SAMPLE EVALUATION

MEAN SQUARED ERROR(MSE)

[7]:
```python
import numpy as np
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Replace '?' with NaN
df.replace('?', np.nan, inplace=True)

# Convert to numeric, coercing invalid values
df = df.apply(pd.to_numeric, errors='coerce')

# Fill missing values with the column mean
df.fillna(df.mean(), inplace=True)

# Features and target variable
X = df[["horsepower", "curb-weight", "engine-size", "highway-mpg"]]
y = df["price"]

# Train a simple linear regression model
model = LinearRegression()
model.fit(X, y)

# Generate predictions
Y_predict_simple_fit = model.predict(X)

# Calculate the Mean Squared Error
mse = mean_squared_error(y, Y_predict_simple_fit)

# Print the MSE value
print(f"Mean Squared Error (MSE): {mse}")
```

```
Mean Squared Error (MSE): 11976801.681229591
```

[9]:
```python
import numpy as np  # For square root calculation

# Given Mean Squared Error (MSE)
mse = 11976801.681229591  # Replace with your actual MSE value

# Calculate the Root Mean Squared Error (RMSE)
rmse = np.sqrt(mse)
```

```
# Print the RMSE value
print(f"Root Mean Squared Error (RMSE): ${rmse:.2f}")
```

```
Root Mean Squared Error (RMSE): $3460.75
```

R-squared/R^2

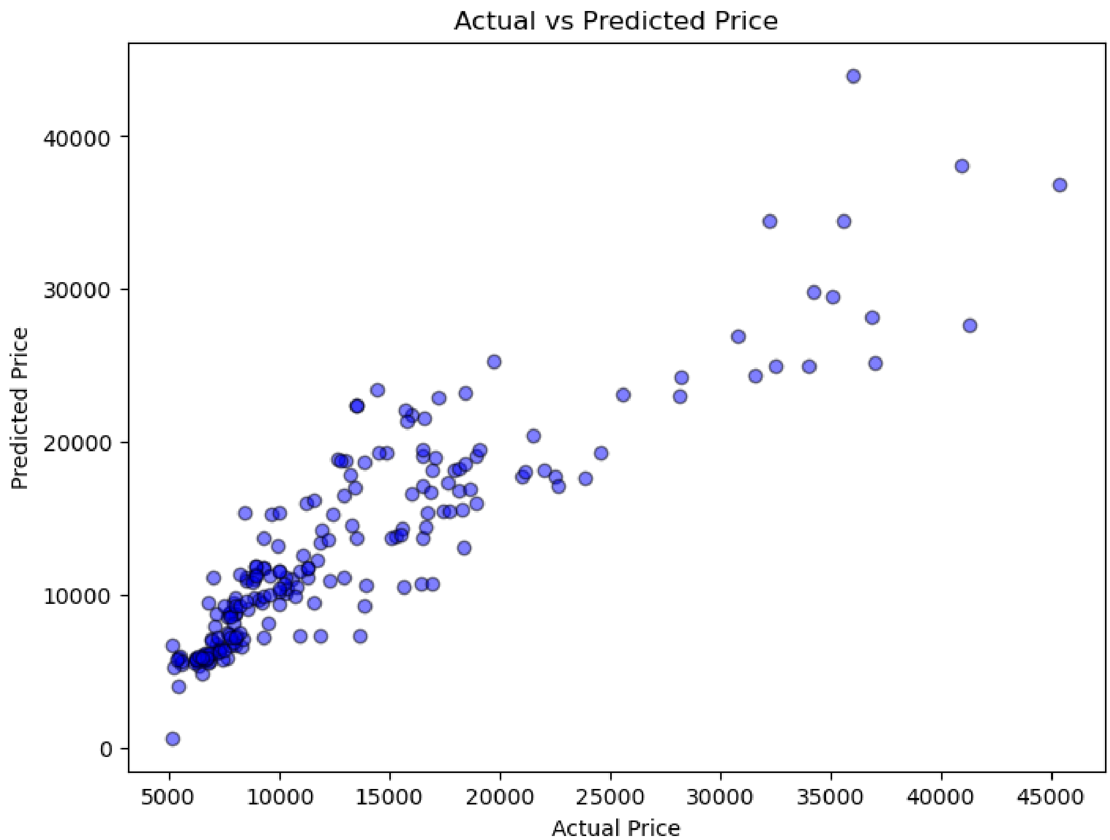
[11]:
```
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

# Assuming 'df' is your DataFrame with 'highway-mpg' and 'price' columns
# Extract the feature and target
X = df[["highway-mpg"]]  # Predictor variable (independent variable)
y = df["price"]          # Target variable (dependent variable)

# Initialize the linear regression model
lm = LinearRegression()

# Fit the model to the data
lm.fit(X, y)

# Predict the target variable using the model
y_pred = lm.predict(X)

# Calculate the R-squared value
r_squared = r2_score(y, y_pred)

# Print the R-squared value
print(f"R-squared (R^2): {r_squared:.4f}")
```

```
R-squared (R^2): 0.4966
```

[13]:
```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt

# Features and target variable
X = df[["horsepower", "curb-weight", "engine-size", "highway-mpg"]]  # Input␣
 ↪features
y = df["price"]  # Target variable

# Initialize the Linear Regression model
lm = LinearRegression()

# Fit the model to the data
lm.fit(X, y)
```

```python
# Predict the target variable (price) using the model
y_pred = lm.predict(X)

# Calculate the R-squared value (R^2)
r_squared = lm.score(X, y)

# Calculate the Mean Squared Error (MSE)
mse = mean_squared_error(y, y_pred)

# Output results
print(f"R-squared (R^2): {r_squared}")
print(f"Mean Squared Error (MSE): {mse}")

# Visualize the actual vs predicted values
plt.figure(figsize=(8, 6))
plt.scatter(y, y_pred, color='blue', edgecolor='k', alpha=0.5)
plt.xlabel('Actual Price')
plt.ylabel('Predicted Price')
plt.title('Actual vs Predicted Price')
plt.show()
```

```
R-squared (R^2): 0.8094130135602673
Mean Squared Error (MSE): 11976801.681229591
```



Hyperparameter Tuning with Grid Search and Random Forest

[19]:
```
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.metrics import mean_squared_error

# Define your feature matrix (X) and target variable (y)
X = df[["horsepower", "curb-weight", "engine-size", "highway-mpg"]]  # Input␣
↪features
y = df["price"]  # Target variable

# Check for missing values
print("Missing values in each column:")
print(df.isnull().sum())

# Check for column names
print("Columns in the dataset:")
print(df.columns)

# Split the data into training and testing sets
```

```python
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,␣
↪random_state=42)

# Check the shape of the train and test sets
print(f"Training data shape: X_train: {X_train.shape}, y_train: {y_train.
↪shape}")
print(f"Testing data shape: X_test: {X_test.shape}, y_test: {y_test.shape}")

# Define the model
rf_model = RandomForestRegressor(random_state=42)

# Define the hyperparameter grid
param_grid = {
    'n_estimators': [100, 200, 300],
    'max_depth': [None, 10, 20, 30],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4],
}

# Perform Grid Search
grid_search = GridSearchCV(estimator=rf_model, param_grid=param_grid, cv=5,␣
↪scoring='neg_mean_squared_error')
print("Starting grid search...")
grid_search.fit(X_train, y_train)

# Get the best parameters
best_params = grid_search.best_params_
print(f"Best Parameters: {best_params}")

# Evaluate the best model
best_rf_model = grid_search.best_estimator_
y_pred = best_rf_model.predict(X_test)

# Evaluate model performance
mse = mean_squared_error(y_test, y_pred)
print(f"Optimized Mean Squared Error: {mse}")
```

```
Missing values in each column:
symboling            201
normalized-losses      0
make                 201
fuel-type            201
aspiration           201
num-of-doors         201
body-style           201
drive-wheels         201
engine-location      201
```

```
wheel-base              0
length                  0
width                   0
height                  0
curb-weight             0
engine-type           201
num-of-cylinders      201
engine-size             0
fuel-system           201
bore                    0
stroke                  0
compression-ratio       0
horsepower              0
peak-rpm                0
city-L/100km            0
highway-mpg             0
price                   0
price-binned          201
dtype: int64
Columns in the dataset:
Index(['symboling', 'normalized-losses', 'make', 'fuel-type', 'aspiration',
       'num-of-doors', 'body-style', 'drive-wheels', 'engine-location',
       'wheel-base', 'length', 'width', 'height', 'curb-weight', 'engine-type',
       'num-of-cylinders', 'engine-size', 'fuel-system', 'bore', 'stroke',
       'compression-ratio', 'horsepower', 'peak-rpm', 'city-L/100km',
       'highway-mpg', 'price', 'price-binned'],
      dtype='object')
Training data shape: X_train: (140, 4), y_train: (140,)
Testing data shape: X_test: (61, 4), y_test: (61,)
Starting grid search…
Best Parameters: {'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split':
2, 'n_estimators': 200}
Optimized Mean Squared Error: 8025825.635245899
```

CONVERT MSE TO USD($)

[21]:
```python
import math

# Optimized Mean Squared Error (MSE)
mse = 8025825.64

# Calculate the root mean squared error (RMSE)
rmse = math.sqrt(mse)

# Print the RMSE in terms of dollar amount
print(f"Root Mean Squared Error (RMSE): ${rmse:.2f}")
```

```
Root Mean Squared Error (RMSE): $2832.99
```

EXTRA CREDIT

[31]:
```python
import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv(r'C:\Users\dj1975\Documents\LinearRegres.csv')

# Verify the DataFrame is loaded
print(df.head())  # Show the first 5 rows to confirm the data
```

```
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3              122.0  alfa-romero       gas        std          two
1        "3              122.0  alfa-romero       gas        std          two
2        "1              122.0  alfa-romero       gas        std          two
3        "2              164.0         audi       gas        std         four
4        "2              164.0         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  fuel-system  \
0  convertible          rwd           front        88.6  …         mpfi
1  convertible          rwd           front        88.6  …         mpfi
2    hatchback          rwd           front        94.5  …         mpfi
3        sedan          fwd           front        99.8  …         mpfi
4        sedan          4wd           front        99.4  …         mpfi

   bore  stroke  compression-ratio horsepower peak-rpm  city-L/100km  \
0  3.47    2.68                9.0        111     5000     11.190476
1  3.47    2.68                9.0        111     5000     11.190476
2  2.68    3.47                9.0        154     5000     12.368421
3  3.19    3.40               10.0        102     5500      9.791667
4  3.19    3.40                8.0        115     5500     13.055556

  highway-mpg  price price-binned
0          27  13495          Low
1          27  16500          Low
2          26  16500          Low
3          30  13950          Low
4          22  17450          Low

[5 rows x 27 columns]
```

[33]:
```python
# Apply one-hot encoding to the 'fuel-type' column
df_encoded = pd.get_dummies(df, columns=["fuel-type"], drop_first=False)
df_encoded[["fuel-type_diesel", "fuel-type_gas"]] =
 ↪df_encoded[["fuel-type_diesel", "fuel-type_gas"]].astype(int)



# Display the first few rows to see the result
#You should see 0 or 1 in "fuel-type_diesel" or "fuel_type_gas"
```

```
print(df_encoded.head())
```

```
  symboling  normalized-losses         make aspiration num-of-doors  \
0        "3              122.0  alfa-romero        std          two
1        "3              122.0  alfa-romero        std          two
2        "1              122.0  alfa-romero        std          two
3        "2              164.0         audi        std         four
4        "2              164.0         audi        std         four

    body-style drive-wheels engine-location  wheel-base    length  …  \
0  convertible          rwd           front        88.6 -0.438315  …
1  convertible          rwd           front        88.6 -0.438315  …
2    hatchback          rwd           front        94.5 -0.243544  …
3        sedan          fwd           front        99.8  0.194690  …
4        sedan          4wd           front        99.4  0.194690  …

   stroke  compression-ratio  horsepower peak-rpm city-L/100km  highway-mpg  \
0    2.68                9.0         111     5000     11.190476           27
1    2.68                9.0         111     5000     11.190476           27
2    3.47                9.0         154     5000     12.368421           26
3    3.40               10.0         102     5500      9.791667           30
4    3.40                8.0         115     5500     13.055556           22

   price price-binned fuel-type_diesel  fuel-type_gas
0  13495          Low                0              1
1  16500          Low                0              1
2  16500          Low                0              1
3  13950          Low                0              1
4  17450          Low                0              1

[5 rows x 28 columns]
```

[35]:
```
# Replace '?' with NaN in the DataFrame
df.replace('?', np.nan, inplace=True)
```

[37]:
```
#Convert Columns to Numeric
df["horsepower"] = pd.to_numeric(df["horsepower"], errors='coerce')
df["curb-weight"] = pd.to_numeric(df["curb-weight"], errors='coerce')
df["engine-size"] = pd.to_numeric(df["engine-size"], errors='coerce')
df["highway-mpg"] = pd.to_numeric(df["highway-mpg"], errors='coerce')
df["price"] = pd.to_numeric(df["price"], errors='coerce')
```

[39]:
```
# Drop rows with missing values
df.dropna(subset=["horsepower", "curb-weight", "engine-size", "highway-mpg",␣
↪"price"], inplace=True)
```

[51]:
```python
from sklearn.ensemble import RandomForestRegressor
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.metrics import mean_squared_error

# Define your feature matrix (X) and target variable (y)
X = df[["horsepower", "curb-weight", "engine-size", "highway-mpg"]]  # Input␣
↪features
y = df["price"]  # Target variable

# Check for missing values
print("Missing values in each column:")
print(df.isnull().sum())

# Check for column names
print("Columns in the dataset:")
print(df.columns)

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,␣
↪random_state=42)

# Check the shape of the train and test sets
print(f"Training data shape: X_train: {X_train.shape}, y_train: {y_train.
↪shape}")
print(f"Testing data shape: X_test: {X_test.shape}, y_test: {y_test.shape}")

# Define the model
rf_model = RandomForestRegressor(random_state=42)

# Define the hyperparameter grid
param_grid = {
    'n_estimators': [100, 200, 300],
    'max_depth': [None, 10, 20, 30],
    'min_samples_split': [2, 5, 10],
    'min_samples_leaf': [1, 2, 4],
}

# Perform Grid Search
grid_search = GridSearchCV(estimator=rf_model, param_grid=param_grid, cv=5,␣
↪scoring='neg_mean_squared_error')
print("Starting grid search...")
grid_search.fit(X_train, y_train)

# Get the best parameters
best_params = grid_search.best_params_
print(f"Best Parameters: {best_params}")
```

```python
# Evaluate the best model
best_rf_model = grid_search.best_estimator_
y_pred = best_rf_model.predict(X_test)

# Evaluate model performance
mse = mean_squared_error(y_test, y_pred)
print(f"Optimized Mean Squared Error: {mse}")
```

```
Missing values in each column:
symboling            0
normalized-losses    0
make                 0
fuel-type            0
aspiration           0
num-of-doors         0
body-style           0
drive-wheels         0
engine-location      0
wheel-base           0
length               0
width                0
height               0
curb-weight          0
engine-type          0
num-of-cylinders     0
engine-size          0
fuel-system          0
bore                 0
stroke               0
compression-ratio    0
horsepower           0
peak-rpm             0
city-L/100km         0
highway-mpg          0
price                0
price-binned         0
dtype: int64
Columns in the dataset:
Index(['symboling', 'normalized-losses', 'make', 'fuel-type', 'aspiration',
       'num-of-doors', 'body-style', 'drive-wheels', 'engine-location',
       'wheel-base', 'length', 'width', 'height', 'curb-weight', 'engine-type',
       'num-of-cylinders', 'engine-size', 'fuel-system', 'bore', 'stroke',
       'compression-ratio', 'horsepower', 'peak-rpm', 'city-L/100km',
       'highway-mpg', 'price', 'price-binned'],
      dtype='object')
Training data shape: X_train: (140, 4), y_train: (140,)
Testing data shape: X_test: (61, 4), y_test: (61,)
```

```
Starting grid search…

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'
```

```
  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
```

```
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
```

```
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:824: UserWarning: Scoring
failed. The score on this train-test partition for these parameters will be set
to nan. Details:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 813, in _score
    scores = scorer(estimator, X_test, y_test)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 266, in __call__
    return self._score(partial(_cached_call, None), estimator, X, y_true,
**_kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/metrics/_scorer.py", line 353, in _score
    y_pred = method_caller(estimator, "predict", X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/metrics/_scorer.py", line 86, in _cached_call
    result, _ = _get_response_values(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_response.py", line 109, in _get_response_values
    y_pred, pos_label = estimator.predict(X), None
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 984, in predict
    X = self._validate_X_predict(X)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 599, in _validate_X_predict
    X = self._validate_data(X, dtype=DTYPE, accept_sparse="csr", reset=False)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 604, in _validate_data
    out = check_array(X, input_name="X", **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py:425: FitFailedWarning:
432 fits failed out of a total of 540.
The score on these train-test partitions for these parameters will be set to
nan.
If these failures are not expected, you can try to debug them by setting
error_score='raise'.

Below are more details about the failures:
--------------------------------------------------------------------------------
432 fits failed with the following error:
Traceback (most recent call last):
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_validation.py", line 732, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py", line 1151, in wrapper
    return fit_method(estimator, *args, **kwargs)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/ensemble/_forest.py", line 348, in fit
    X, y = self._validate_data(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
```

```
packages/sklearn/base.py", line 621, in _validate_data
    X, y = check_X_y(X, y, **check_params)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 1147, in check_X_y
    X = check_array(
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/validation.py", line 917, in check_array
    array = _asarray_with_order(array, order=order, dtype=dtype, xp=xp)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/utils/_array_api.py", line 380, in _asarray_with_order
    array = numpy.asarray(array, order=order, dtype=dtype)
  File "/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/pandas/core/generic.py", line 2084, in __array__
    arr = np.asarray(values, dtype=dtype)
ValueError: could not convert string to float: '?'

  warnings.warn(some_fits_failed_message, FitFailedWarning)
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/model_selection/_search.py:976: UserWarning: One or more of the
test scores are non-finite: [nan nan nan nan nan nan nan nan nan nan nan nan nan
nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan
 nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan nan]
  warnings.warn(
```

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
/tmp/ipykernel_143/3573138932.py in ?()
     34
     35 # Perform Grid Search
     36 grid_search = GridSearchCV(estimator=rf_model, param_grid=param_grid,␣
 ↪cv=5, scoring='neg_mean_squared_error')
     37 print("Starting grid search…")
---> 38 grid_search.fit(X_train, y_train)
     39
     40 # Get the best parameters
     41 best_params = grid_search.best_params_

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/bas .
 ↪py in ?(estimator, *args, **kwargs)
   1147                 skip_parameter_validation=(
   1148                     prefer_skip_nested_validation or␣
 ↪global_skip_validation
   1149                 )
```

```
   1150             ):
-> 1151                 return fit_method(estimator, *args, **kwargs)

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
 ↪model_selection/_search.py in ?(self, X, y, groups, **fit_params)
    929                 clone(base_estimator).set_params(**self.best_params_)
    930             )
    931             refit_start_time = time.time()
    932             if y is not None:
--> 933                 self.best_estimator_.fit(X, y, **fit_params)
    934             else:
    935                 self.best_estimator_.fit(X, **fit_params)
    936             refit_end_time = time.time()

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/bas‹.
 ↪py in ?(estimator, *args, **kwargs)
   1147                 skip_parameter_validation=(
   1148                     prefer_skip_nested_validation or␣
 ↪global_skip_validation
   1149                 )
   1150             ):
-> 1151                 return fit_method(estimator, *args, **kwargs)

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
 ↪ensemble/_forest.py in ?(self, X, y, sample_weight)
    344         """
    345         # Validate or convert input data
    346         if issparse(y):
    347             raise ValueError("sparse multilabel-indicator for y is not␣
 ↪supported.")
--> 348         X, y = self._validate_data(

    349             X, y, multi_output=True, accept_sparse="csc", dtype=DTYPE
    350         )
    351         if sample_weight is not None:

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/bas‹.
 ↪py in ?(self, X, y, reset, validate_separately, cast_to_ndarray,␣
 ↪**check_params)
    617                 if "estimator" not in check_y_params:
    618                     check_y_params = {**default_check_params,␣
 ↪**check_y_params}
    619                 y = check_array(y, input_name="y", **check_y_params)
    620             else:
--> 621                 X, y = check_X_y(X, y, **check_params)
    622             out = X, y
    623 
    624         if not no_val_X and check_params.get("ensure_2d", True):
```

```
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
↪utils/validation.py in ?(X, y, accept_sparse, accept_large_sparse, dtype,␣
↪order, copy, force_all_finite, ensure_2d, allow_nd, multi_output,␣
↪ensure_min_samples, ensure_min_features, y_numeric, estimator)
   1143         raise ValueError(
   1144             f"{estimator_name} requires y to be passed, but the target ␣
↪is None"
   1145         )
   1146
-> 1147     X = check_array(
   1148         X,
   1149         accept_sparse=accept_sparse,
   1150         accept_large_sparse=accept_large_sparse,

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
↪utils/validation.py in ?(array, accept_sparse, accept_large_sparse, dtype,␣
↪order, copy, force_all_finite, ensure_2d, allow_nd, ensure_min_samples,␣
↪ensure_min_features, estimator, input_name)
    914                     )
    915                 array = xp.astype(array, dtype, copy=False)
    916             else:
    917                 array = _asarray_with_order(array, order=order,␣
↪dtype=dtype, xp=xp)
--> 918         except ComplexWarning as complex_warning:
    919             raise ValueError(
    920                 "Complex data not supported\n{}\n".format(array)
    921             ) from complex_warning

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/sklearn/
↪utils/_array_api.py in ?(array, dtype, order, copy, xp)
    376         # Use NumPy API to support order
    377         if copy is True:
    378             array = numpy.array(array, order=order, dtype=dtype)
    379         else:
--> 380             array = numpy.asarray(array, order=order, dtype=dtype)
    381
    382         # At this point array is a NumPy ndarray. We convert it to an␣
↪array
    383         # container that is consistent with the input's namespace.

/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-packages/pandas/core/
↪generic.py in ?(self, dtype)
   2082     def __array__(self, dtype: npt.DTypeLike | None = None) -> np.
↪ndarray:
   2083         values = self._values
-> 2084         arr = np.asarray(values, dtype=dtype)
   2085         if (
```

```
   2086                 astype_is_view(values.dtype, arr.dtype)
   2087                 and using_copy_on_write()

ValueError: could not convert string to float: '?'
```

PREDICTION AND DECISION MAKING

DOES THE PREDICTED VALUE MAKE SENSE?

[59]:
```
# Assuming lm is your LinearRegression model
lm.fit(df['highway-mpg'].values.reshape(-1, 1), df['price'])

# Now you can make a prediction for a highway-mpg value of 30.0
prediction = lm.predict(np.array(30.0).reshape(-1, 1))
print(prediction)
```

```
[13771.3045085]
```

The model predicts a price of approximately 13,771.30 for a vehicle with a highway mileage of 30.0 miles per gallon.

[61]:
```
from sklearn.linear_model import LinearRegression
import numpy as np

# Assuming df is your DataFrame
lm = LinearRegression()

# Fit the model (make sure you reshape 'highway-mpg' to 2D array for sklearn)
lm.fit(df['highway-mpg'].values.reshape(-1, 1), df['price'])

# Make a prediction for a highway-mpg value of 30.0
prediction = lm.predict(np.array(30.0).reshape(-1, 1))

# Print the prediction
print("Predicted price:", prediction)

# Get the model's coefficient (slope of the line)
print("Model coefficient (slope):", lm.coef_)
```

```
Predicted price: [13771.3045085]
Model coefficient (slope): [-821.73337832]
```

1)Predicted price: The model predicts that a car with a highway MPG of 30.0 will have a price of approximately $13,771.30.

2)Model coefficient (slope): The coefficient of -821.73 indicates that for each additional unit increase in highway-mpg, the predicted price of the car decreases by $821.73. This suggests an inverse relationship between highway MPG and the car's price.

In simpler terms, cars with higher highway MPG tend to be less expensive based on the data you're working with. This could indicate that cars with better fuel efficiency might be older or have less expensive features, or it could reflect the general trend in the dataset you have.

[65]:
```python
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression

# Assuming 'df' is your DataFrame containing the data

# Clean the data: Replace '?' with NaN, then convert columns to numeric
df['highway-mpg'] = pd.to_numeric(df['highway-mpg'], errors='coerce')  #␣
↪Converts invalid entries to NaN
df['horsepower'] = pd.to_numeric(df['horsepower'], errors='coerce')  # Converts␣
↪invalid entries to NaN

# Drop rows with missing values in 'highway-mpg' or 'horsepower'
df = df.dropna(subset=['highway-mpg', 'horsepower', 'price'])

# Select the features and target
X = df[['highway-mpg', 'horsepower']]  # Use both 'highway-mpg' and 'horsepower'
y = df['price']

# Initialize the model
lm = LinearRegression()

# Train the model with the selected features
lm.fit(X, y)

# Predict the price for a car with highway-mpg of 30 and horsepower of 130
predicted_price = lm.predict(np.array([[30.0, 130]]))  # Input values:␣
↪[highway-mpg, horsepower]

# Get the model's coefficients (slopes)
model_coefficients = lm.coef_

# Output the predicted price and coefficients
print(f"Predicted price: {predicted_price[0]}")
print(f"Model coefficients (slopes): {model_coefficients}")
```

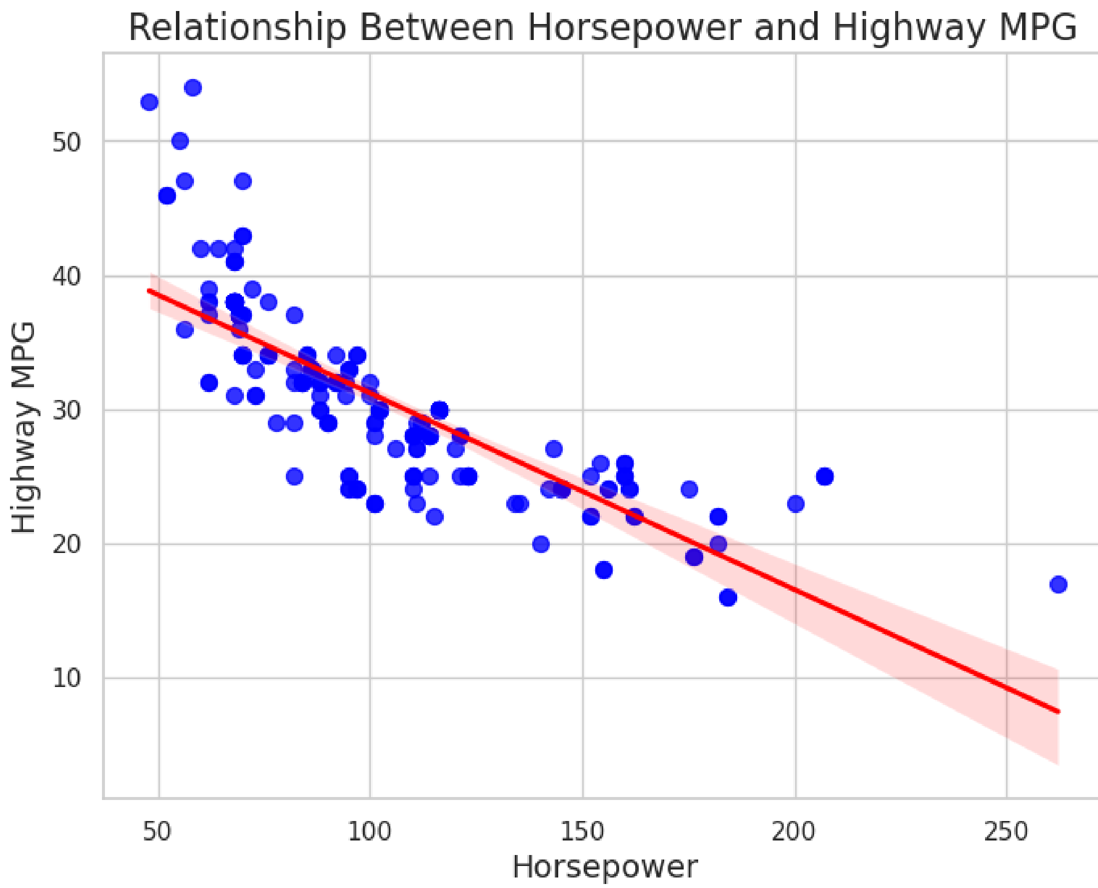
```
Predicted price: 17259.94325654814
Model coefficients (slopes): [-175.34029281  146.475147  ]
```

```
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/sklearn/base.py:464: UserWarning: X does not have valid feature names,
but LinearRegression was fitted with feature names
  warnings.warn(
```

#The coefficient for highway-mpg is -175.34. This means that for each 1-unit increase in highway-mpg, the price of the car decreases by approximately $175.34. #The coefficient for horsepower is

146.48. This indicates that for each 1-unit increase in horsepower, the price of the car increases by approximately $146.48. #The model suggests that higher horsepower increases the car's price, while higher highway-mpg slightly decreases the price.

[67]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load your data (replace with your actual DataFrame)
# df = pd.read_csv("your_data.csv")

# Example of how the DataFrame should look
# Assuming you already have the df with columns 'horsepower' and 'highway-mpg'
# df = pd.DataFrame({'horsepower': [100, 150, 200, 250, 300], 'highway-mpg':␣
↪[30, 28, 25, 22, 20]})

# Create a scatter plot and a regression line
sns.set(style="whitegrid")
plt.figure(figsize=(8, 6))
sns.regplot(x='horsepower', y='highway-mpg', data=df, scatter_kws={'s': 50,␣
↪'color': 'blue'}, line_kws={'color': 'red'})

# Set labels and title
plt.title('Relationship Between Horsepower and Highway MPG', fontsize=16)
plt.xlabel('Horsepower', fontsize=14)
plt.ylabel('Highway MPG', fontsize=14)

# Show plot
plt.show()

# Calculate and display correlation coefficient
correlation = df['horsepower'].corr(df['highway-mpg'])
print(f'Correlation between horsepower and highway-mpg: {correlation:.2f}')
```



Correlation between horsepower and highway-mpg: -0.80

PLOT "HIGHWAY-MPG" IN A RANGE FROM (0-100)

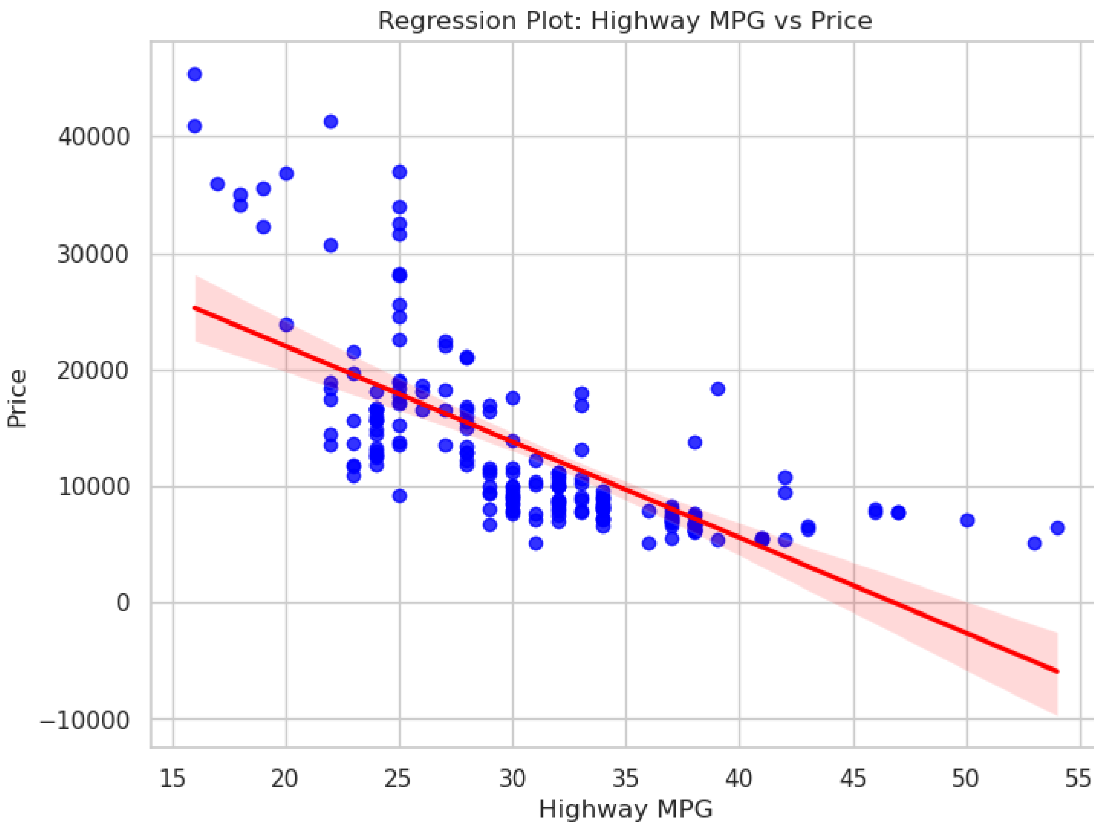
[89]:
```
#Generate a sequence of values in a specified range using Numpy
import numpy as np

# Generate new input values for 'highway-mpg' (from 1 to 100)
highway_mpg_input = np.arange(1, 101, 1).reshape(-1, 1)  # Highway-mpg values␣
    ↪from 1 to 100

# Make predictions using the trained model (only using highway-mpg)
Yhat = lm.predict(highway_mpg_input)

# Display the predictions
print(Yhat)
```

```
[ 37628.49946606  36806.99474391  35985.49002176  35163.98529961
  34342.48057746  33520.9758553   32699.47113315  31877.966411
  31056.46168885  30234.9569667   29413.45224455  28591.9475224
```

```
 27770.44280025  26948.9380781   26127.43335595  25305.92863379
 24484.42391164  23662.91918949  22841.41446734  22019.90974519
 21198.40502304  20376.90030089  19555.39557874  18733.89085659
 17912.38613444  17090.88141229  16269.37669013  15447.87196798
 14626.36724583  13804.86252368  12983.35780153  12161.85307938
 11340.34835723  10518.84363508   9697.33891293   8875.83419078
  8054.32946862   7232.82474647   6411.32002432   5589.81530217
  4768.31058002   3946.80585787   3125.30113572   2303.79641357
  1482.29169142    660.78696927   -160.71775289   -982.22247504
 -1803.72719719  -2625.23191934  -3446.73664149  -4268.24136364
 -5089.74608579  -5911.25080794  -6732.75553009  -7554.26025224
 -8375.7649744   -9197.26969655 -10018.7744187  -10840.27914085
-11661.783863   -12483.28858515 -13304.7933073  -14126.29802945
-14947.8027516  -15769.30747375 -16590.81219591 -17412.31691806
-18233.82164021 -19055.32636236 -19876.83108451 -20698.33580666
-21519.84052881 -22341.34525096 -23162.84997311 -23984.35469526
-24805.85941742 -25627.36413957 -26448.86886172 -27270.37358387
-28091.87830602 -28913.38302817 -29734.88775032 -30556.39247247
-31377.89719462 -32199.40191677 -33020.90663893 -33842.41136108
-34663.91608323 -35485.42080538 -36306.92552753 -37128.43024968
-37949.93497183 -38771.43969398 -39592.94441613 -40414.44913828
-41235.95386044 -42057.45858259 -42878.96330474 -43700.46802689]
```

VISUALIZE THE RESULTS

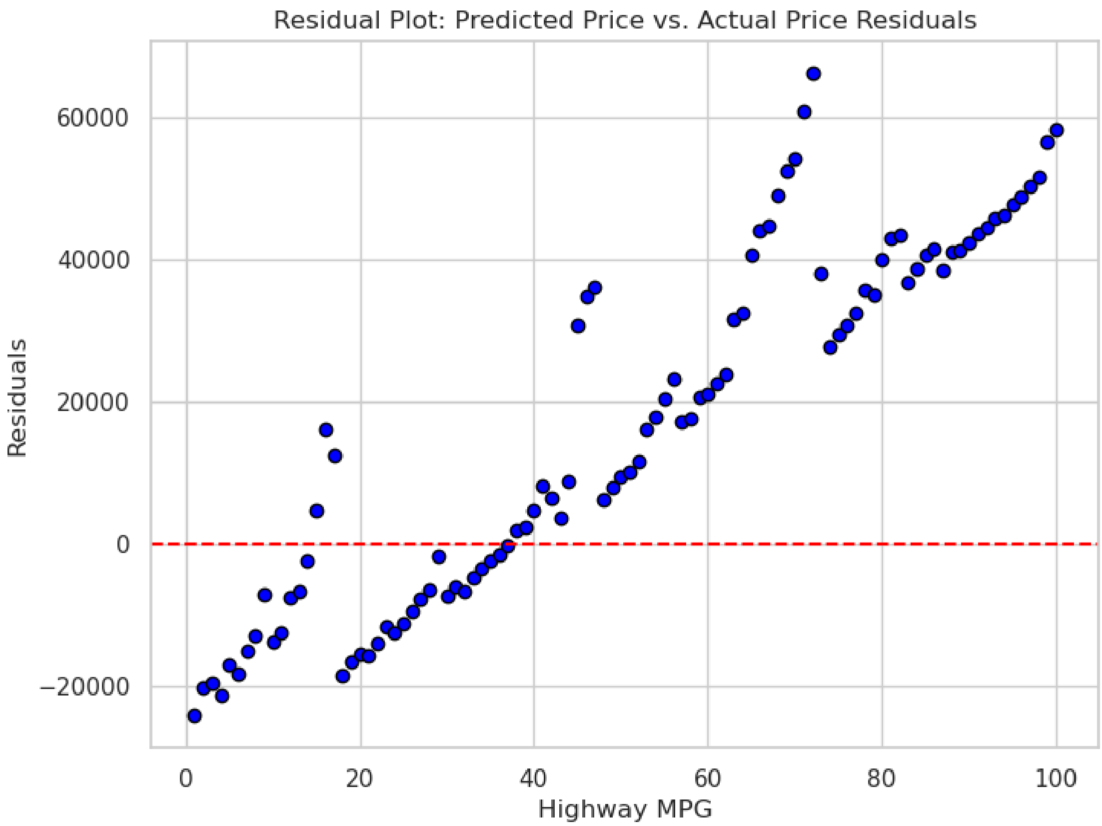
[91]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Assuming 'df' contains the columns 'price' and 'highway-mpg'

# Create a regression plot (highway-mpg vs price)
plt.figure(figsize=(8, 6))
sns.regplot(x='highway-mpg', y='price', data=df, scatter_kws={'color':'blue'},
 ↪line_kws={'color':'red'})

# Title and labels
plt.title('Regression Plot: Highway MPG vs Price')
plt.xlabel('Highway MPG')
plt.ylabel('Price')

# Show the plot
plt.show()
```



[93]:

```
import numpy as np
import matplotlib.pyplot as plt

# Generate new input values for 'highway-mpg' (from 1 to 100)
highway_mpg_input = np.arange(1, 101, 1).reshape(-1, 1)

# Make predictions using the trained model (only using highway-mpg)
Yhat = lm.predict(highway_mpg_input)

# Calculate residuals (difference between actual and predicted values)
# Assuming y_actual contains the true price values corresponding to the input␣
↪data.
# For illustration, let's assume y_actual is the actual price for each␣
↪highway-mpg value.
y_actual = df['price'].values[:100]  # Adjust depending on the available data

# Calculate residuals
residuals = y_actual - Yhat

# Plot the residuals
```

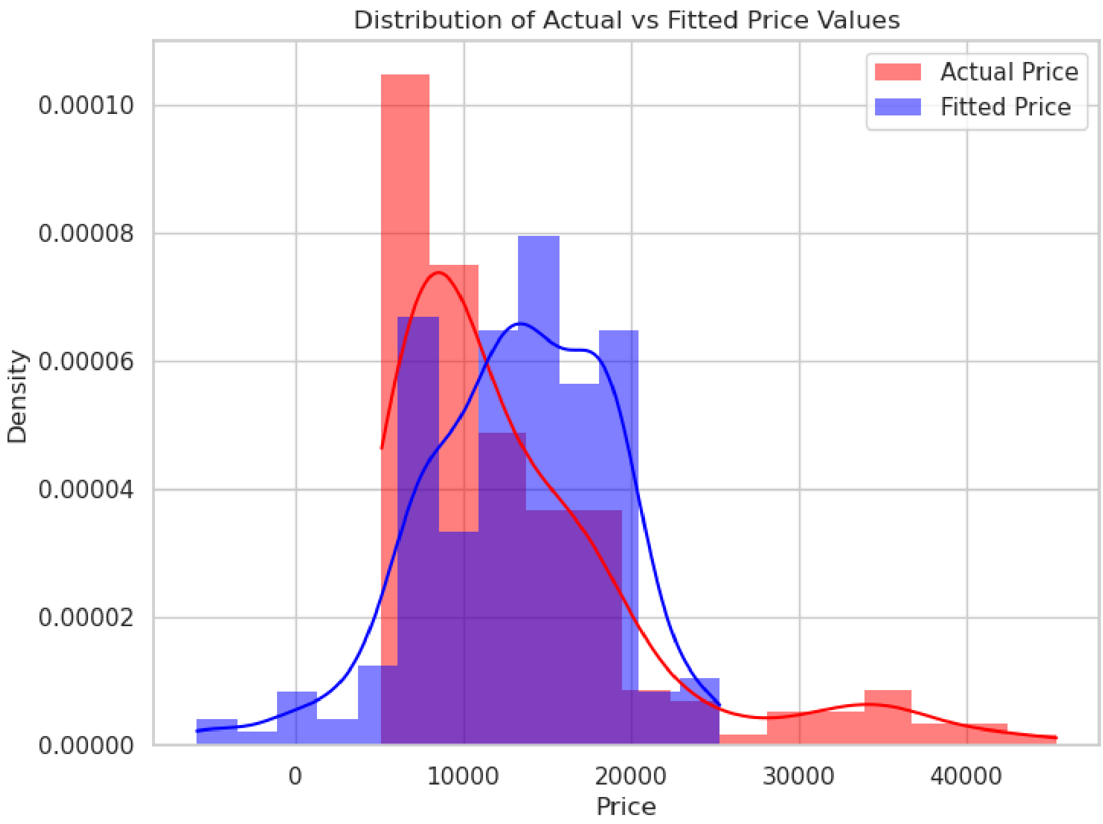
```
plt.figure(figsize=(8, 6))
plt.scatter(highway_mpg_input, residuals, color='blue', edgecolor='black')
plt.axhline(y=0, color='red', linestyle='--')
plt.title('Residual Plot: Predicted Price vs. Actual Price Residuals')
plt.xlabel('Highway MPG')
plt.ylabel('Residuals')
plt.show()
```



[99]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Assuming 'y' is your actual target values (price)
# 'Yhat' is the predicted (fitted) values from the model

plt.figure(figsize=(8, 6))

# Plot the distribution of actual values (red) and fitted values (blue)
sns.histplot(y, color='red', kde=True, label='Actual Price', stat='density',␣
 ↪linewidth=0)
```

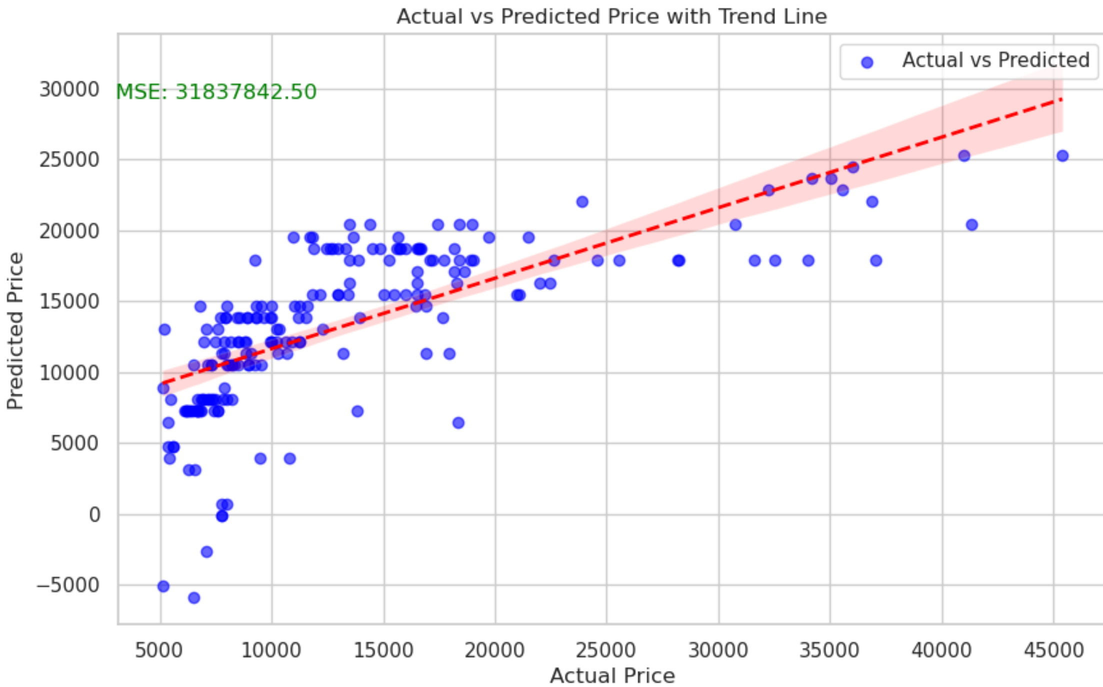
```python
sns.histplot(Yhat, color='blue', kde=True, label='Fitted Price',␣
↪stat='density', linewidth=0)

# Add labels and title
plt.title('Distribution of Actual vs Fitted Price Values')
plt.xlabel('Price')
plt.ylabel('Density')

# Show the legend
plt.legend()

# Display the plot
plt.show()
```

```
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/seaborn/_oldcore.py:1119: FutureWarning: use_inf_as_na option is
deprecated and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
/opt/conda/envs/anaconda-2024.02-py310/lib/python3.10/site-
packages/seaborn/_oldcore.py:1119: FutureWarning: use_inf_as_na option is
deprecated and will be removed in a future version. Convert inf values to NaN
before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```



[101]:
```
import matplotlib.pyplot as plt
import numpy as np
from sklearn.metrics import mean_squared_error
import seaborn as sns

# Assuming 'y' is the actual target values (price)
# 'Yhat' is the predicted (fitted) values from the model

# Calculate the Mean Squared Error
mse = mean_squared_error(y, Yhat)

# Plot the actual vs predicted values (fitted values)
plt.figure(figsize=(10, 6))
plt.scatter(y, Yhat, color='blue', label='Actual vs Predicted', alpha=0.6)

# Add a trend line (Line of Best Fit) to the plot
sns.regplot(x=y, y=Yhat, scatter=False, color='red', line_kws={'lw': 2, 'ls':␣
↪'--'})

# Set labels and title
plt.xlabel('Actual Price')
```

```
plt.ylabel('Predicted Price')
plt.title('Actual vs Predicted Price with Trend Line')

# Display MSE on the plot
plt.text(0.1, 0.9, f'MSE: {mse:.2f}', ha='center', va='center', transform=plt.
  ↪gca().transAxes, fontsize=12, color='green')

# Show the plot
plt.legend()
plt.show()
```



[105]:
```
import numpy as np

# Assuming 'y' is the actual values and 'Yhat' is the predicted values
mse = np.mean((y - Yhat) ** 2)  # Calculate MSE
rmse = np.sqrt(mse)  # Calculate RMSE

# Display the RMSE value
print(f"RMSE: {rmse}")
```

```
RMSE: 5642.503212162324
```

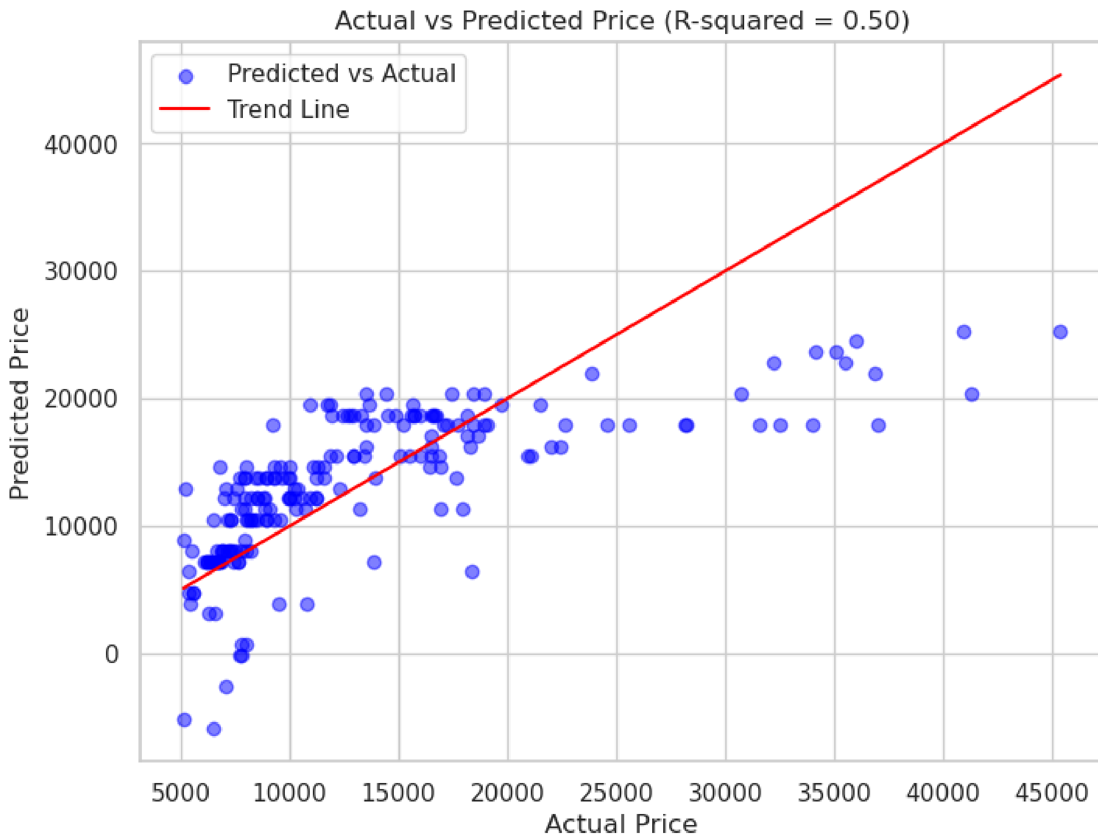
[107]:
```
from sklearn.metrics import r2_score
```

```python
# Assuming 'y' is the actual values and 'Yhat' is the predicted values
r2 = r2_score(y, Yhat)

# Display the R-squared value
print(f"R-squared: {r2}")
```

R-squared: 0.4973491560296689

[109]:
```python
import numpy as np
import matplotlib.pyplot as plt
from sklearn.metrics import r2_score
from sklearn.linear_model import LinearRegression

# Assuming 'X' is the feature matrix (e.g., highway-mpg) and 'y' is the target␣
↪variable (price)
# Fit the model
lm = LinearRegression()
lm.fit(X, y)

# Make predictions
Yhat = lm.predict(X)

# Calculate R-squared
r2 = r2_score(y, Yhat)

# Print the R-squared value
print(f"R-squared: {r2}")

# Visualization of actual vs. predicted values
plt.figure(figsize=(8, 6))

# Scatter plot of actual values vs predicted values
plt.scatter(y, Yhat, color='blue', alpha=0.5, label='Predicted vs Actual')

# Plotting the trend line
plt.plot([min(y), max(y)], [min(y), max(y)], color='red', label='Trend Line')

plt.title(f'Actual vs Predicted Price (R-squared = {r2:.2f})')
plt.xlabel('Actual Price')
plt.ylabel('Predicted Price')
plt.legend()

plt.show()
```

R-squared: 0.4973491560296689



[1]:
```python
import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv(r'C:\Users\dj1975\Documents\LinearRegres.csv')

# Verify the DataFrame is loaded
print(df.head())  # Show the first 5 rows to confirm the data
```

```
  symboling  normalized-losses         make fuel-type aspiration num-of-doors  \
0        "3              122.0  alfa-romero       gas        std          two
1        "3              122.0  alfa-romero       gas        std          two
2        "1              122.0  alfa-romero       gas        std          two
3        "2              164.0         audi       gas        std         four
4        "2              164.0         audi       gas        std         four

    body-style drive-wheels engine-location  wheel-base  …  fuel-system  \
0  convertible          rwd           front        88.6  …         mpfi
1  convertible          rwd           front        88.6  …         mpfi
2    hatchback          rwd           front        94.5  …         mpfi
3        sedan          fwd           front        99.8  …         mpfi
4        sedan          4wd           front        99.4  …         mpfi
```

```
   bore  stroke  compression-ratio horsepower peak-rpm  city-L/100km  \
0  3.47    2.68                9.0         111     5000     11.190476
1  3.47    2.68                9.0         111     5000     11.190476
2  2.68    3.47                9.0         154     5000     12.368421
3  3.19    3.40               10.0         102     5500      9.791667
4  3.19    3.40                8.0         115     5500     13.055556

  highway-mpg  price price-binned
0          27  13495          Low
1          27  16500          Low
2          26  16500          Low
3          30  13950          Low
4          22  17450          Low

[5 rows x 27 columns]
```

MODEL EVALUATION AND REFINEMENT

FUNCTION TRAINING TEST SPLIT()

[3]:
```python
from sklearn.model_selection import train_test_split

# Assuming 'df' is your dataset and 'price' is the target variable
x_data = df.drop(columns=['price'])  # Drop the target column from the feature␣
 ↪set
y_data = df['price']  # Target variable (price)

# Splitting the data into training and testing sets
x_train, x_test, y_train, y_test = train_test_split(x_data, y_data, test_size=0.
 ↪3, random_state=0)

# Optionally, print the shapes of the splits to verify
print(f"Training data (features): {x_train.shape}")
print(f"Testing data (features): {x_test.shape}")
print(f"Training data (target): {y_train.shape}")
print(f"Testing data (target): {y_test.shape}")
```

```
Training data (features): (140, 26)
Testing data (features): (61, 26)
Training data (target): (140,)
Testing data (target): (61,)
```

[55]:
```python
import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv(r'LinearRegres.csv')
```

```python
# Verify the DataFrame is loaded
print(df.head(25))  # Show the first 25 rows to confirm the data
print(df.tail(25))  #Show the last 25 rows to confirm the data
```

```
   symboling  normalized-losses         make fuel-type aspiration  \
0         "3              122.0  alfa-romero       gas        std
1         "3              122.0  alfa-romero       gas        std
2         "1              122.0  alfa-romero       gas        std
3         "2              164.0         audi       gas        std
4         "2              164.0         audi       gas        std
5         "2              122.0         audi       gas        std
6         "1              158.0         audi       gas        std
7         "1              122.0         audi       gas        std
8         "1              158.0         audi       gas      turbo
9         "2              192.0          bmw       gas        std
10        "0              192.0          bmw       gas        std
11        "0              188.0          bmw       gas        std
12        "0              188.0          bmw       gas        std
13        "1              122.0          bmw       gas        std
14        "0              122.0          bmw       gas        std
15        "0              122.0          bmw       gas        std
16        "0              122.0          bmw       gas        std
17        "2              121.0    chevrolet       gas        std
18        "1               98.0    chevrolet       gas        std
19        "0               81.0    chevrolet       gas        std
20        "1              118.0        dodge       gas        std
21        "1              118.0        dodge       gas        std
22        "1              118.0        dodge       gas      turbo
23        "1              148.0        dodge       gas        std
24        "1              148.0        dodge       gas        std

   num-of-doors   body-style drive-wheels engine-location  wheel-base  …  \
0           two  convertible          rwd           front        88.6  …
1           two  convertible          rwd           front        88.6  …
2           two    hatchback          rwd           front        94.5  …
3          four        sedan          fwd           front        99.8  …
4          four        sedan          4wd           front        99.4  …
5           two        sedan          fwd           front        99.8  …
6          four        sedan          fwd           front       105.8  …
7          four        wagon          fwd           front       105.8  …
8          four        sedan          fwd           front       105.8  …
9           two        sedan          rwd           front       101.2  …
10         four        sedan          rwd           front       101.2  …
11          two        sedan          rwd           front       101.2  …
12         four        sedan          rwd           front       101.2  …
13         four        sedan          rwd           front       103.5  …
14         four        sedan          rwd           front       103.5  …
```

```
15          two       sedan         rwd          front       103.5  …
16         four       sedan         rwd          front       110.0  …
17          two   hatchback         fwd          front        88.4  …
18          two   hatchback         fwd          front        94.5  …
19         four       sedan         fwd          front        94.5  …
20          two   hatchback         fwd          front        93.7  …
21          two   hatchback         fwd          front        93.7  …
22          two   hatchback         fwd          front        93.7  …
23         four   hatchback         fwd          front        93.7  …
24         four       sedan         fwd          front        93.7  …

   fuel-system  bore  stroke  compression-ratio horsepower peak-rpm  \
0         mpfi  3.47    2.68               9.00        111     5000
1         mpfi  3.47    2.68               9.00        111     5000
2         mpfi  2.68    3.47               9.00        154     5000
3         mpfi  3.19    3.40              10.00        102     5500
4         mpfi  3.19    3.40               8.00        115     5500
5         mpfi  3.19    3.40               8.50        110     5500
6         mpfi  3.19    3.40               8.50        110     5500
7         mpfi  3.19    3.40               8.50        110     5500
8         mpfi  3.13    3.40               8.30        140     5500
9         mpfi  3.50    2.80               8.80        101     5800
10        mpfi  3.50    2.80               8.80        101     5800
11        mpfi  3.31    3.19               9.00        121     4250
12        mpfi  3.31    3.19               9.00        121     4250
13        mpfi  3.31    3.19               9.00        121     4250
14        mpfi  3.62    3.39               8.00        182     5400
15        mpfi  3.62    3.39               8.00        182     5400
16        mpfi  3.62    3.39               8.00        182     5400
17        2bbl  2.91    3.03               9.50         48     5100
18        2bbl  3.03    3.11               9.60         70     5400
19        2bbl  3.03    3.11               9.60         70     5400
20        2bbl  2.97    3.23               9.41         68     5500
21        2bbl  2.97    3.23               9.40         68     5500
22        mpfi  3.03    3.39               7.60        102     5500
23        2bbl  2.97    3.23               9.40         68     5500
24        2bbl  2.97    3.23               9.40         68     5500

   city-L/100km highway-mpg  price price-binned
0     11.190476          27  13495          Low
1     11.190476          27  16500          Low
2     12.368421          26  16500          Low
3      9.791667          30  13950          Low
4     13.055556          22  17450          Low
5     12.368421          25  15250          Low
6     12.368421          25  17710          Low
7     12.368421          25  18920       Medium
8     13.823529          20  23875       Medium
```

```
9     10.217391          29  16430         Low
10    10.217391          29  16925         Low
11    11.190476          28  20970      Medium
12    11.190476          28  21105      Medium
13    11.750000          25  24565      Medium
14    14.687500          22  30760      Medium
15    14.687500          22  41315        High
16    15.666667          20  36880        High
17     5.000000          53   5151         Low
18     6.184211          43   6295         Low
19     6.184211          43   6575         Low
20     6.351351          41   5572         Low
21     7.580645          38   6377         Low
22     9.791667          30   7957         Low
23     7.580645          38   6229         Low
24     7.580645          38   6692         Low

[25 rows x 27 columns]
    symboling  normalized-losses        make fuel-type aspiration  \
176       "-1               90.0      toyota       gas        std
177       "-1              122.0      toyota       gas        std
178        "2              122.0  volkswagen    diesel        std
179        "2              122.0  volkswagen       gas        std
180        "2               94.0  volkswagen    diesel        std
181        "2               94.0  volkswagen       gas        std
182        "2               94.0  volkswagen       gas        std
183        "2               94.0  volkswagen    diesel      turbo
184        "2               94.0  volkswagen       gas        std
185        "3              122.0  volkswagen       gas        std
186        "3              256.0  volkswagen       gas        std
187        "0              122.0  volkswagen       gas        std
188        "0              122.0  volkswagen    diesel      turbo
189        "0              122.0  volkswagen       gas        std
190       "-2              103.0       volvo       gas        std
191       "-1               74.0       volvo       gas        std
192       "-2              103.0       volvo       gas        std
193       "-1               74.0       volvo       gas        std
194       "-2              103.0       volvo       gas      turbo
195       "-1               74.0       volvo       gas      turbo
196       "-1               95.0       volvo       gas        std
197       "-1               95.0       volvo       gas      turbo
198       "-1               95.0       volvo       gas        std
199       "-1               95.0       volvo    diesel      turbo
200       "-1               95.0       volvo       gas      turbo

    num-of-doors   body-style drive-wheels engine-location  wheel-base  …  \
176         four        sedan          rwd           front       104.5  …
177         four        wagon          rwd           front       104.5  …
```

178	two	sedan	fwd	front	97.3	…
179	two	sedan	fwd	front	97.3	…
180	four	sedan	fwd	front	97.3	…
181	four	sedan	fwd	front	97.3	…
182	four	sedan	fwd	front	97.3	…
183	four	sedan	fwd	front	97.3	…
184	four	sedan	fwd	front	97.3	…
185	two	convertible	fwd	front	94.5	…
186	two	hatchback	fwd	front	94.5	…
187	four	sedan	fwd	front	100.4	…
188	four	sedan	fwd	front	100.4	…
189	four	wagon	fwd	front	100.4	…
190	four	sedan	rwd	front	104.3	…
191	four	wagon	rwd	front	104.3	…
192	four	sedan	rwd	front	104.3	…
193	four	wagon	rwd	front	104.3	…
194	four	sedan	rwd	front	104.3	…
195	four	wagon	rwd	front	104.3	…
196	four	sedan	rwd	front	109.1	…
197	four	sedan	rwd	front	109.1	…
198	four	sedan	rwd	front	109.1	…
199	four	sedan	rwd	front	109.1	…
200	four	sedan	rwd	front	109.1	…

	fuel-system	bore	stroke	compression-ratio	horsepower	peak-rpm	\
176	mpfi	3.27	3.35	9.2	156	5200	
177	mpfi	3.27	3.35	9.2	156	5200	
178	idi	3.01	3.40	23.0	52	4800	
179	mpfi	3.19	3.40	9.0	85	5250	
180	idi	3.01	3.40	23.0	52	4800	
181	mpfi	3.19	3.40	9.0	85	5250	
182	mpfi	3.19	3.40	9.0	85	5250	
183	idi	3.01	3.40	23.0	68	4500	
184	mpfi	3.19	3.40	10.0	100	5500	
185	mpfi	3.19	3.40	8.5	90	5500	
186	mpfi	3.19	3.40	8.5	90	5500	
187	mpfi	3.19	3.40	8.5	110	5500	
188	idi	3.01	3.40	23.0	68	4500	
189	mpfi	3.19	3.40	9.0	88	5500	
190	mpfi	3.78	3.15	9.5	114	5400	
191	mpfi	3.78	3.15	9.5	114	5400	
192	mpfi	3.78	3.15	9.5	114	5400	
193	mpfi	3.78	3.15	9.5	114	5400	
194	mpfi	3.62	3.15	7.5	162	5100	
195	mpfi	3.62	3.15	7.5	162	5100	
196	mpfi	3.78	3.15	9.5	114	5400	
197	mpfi	3.78	3.15	8.7	160	5300	
198	mpfi	3.58	2.87	8.8	134	5500	

```
199          idi  3.01    3.40               23.0         106     4800
200         mpfi  3.78    3.15                9.5         114     5400

     city-L/100km highway-mpg  price price-binned
176     11.750000          24  15690          Low
177     12.368421          24  15750          Low
178      6.351351          46   7775          Low
179      8.703704          34   7975          Low
180      6.351351          46   7995          Low
181      8.703704          34   8195          Low
182      8.703704          34   8495          Low
183      6.351351          42   9495          Low
184      9.038462          32   9995          Low
185      9.791667          29  11595          Low
186      9.791667          29   9980          Low
187     12.368421          24  13295          Low
188      7.121212          38  13845          Low
189      9.400000          31  12290          Low
190     10.217391          28  12940          Low
191     10.217391          28  13415          Low
192      9.791667          28  15985          Low
193      9.791667          28  16515          Low
194     13.823529          22  18420          Low
195     13.823529          22  18950       Medium
196     10.217391          28  16845          Low
197     12.368421          25  19045       Medium
198     13.055556          23  21485       Medium
199      9.038462          27  22470       Medium
200     12.368421          25  22625       Medium

[25 rows x 27 columns]
```

[ ]: