

PRACTICAL-5

NAME-PRIYANSH PANCHAL

ROLLNO-24BEE106

```
import pandas as pd
import numpy as np

data = pd.read_csv("BostonHousing.csv", header='infer')

X = data.iloc[:, :-1] # All columns except the last one
y = data.iloc[:, -1] # Last column

print("Total Rows:", data.shape[0])

test_split = 0.1

indices = np.random.permutation(data.shape[0])
n_train = int((1 - test_split) * data.shape[0])

X_train = X.iloc[indices[:n_train]]
y_train = y.iloc[indices[:n_train]]

X_test = X.iloc[indices[n_train:]]
y_test = y.iloc[indices[n_train:]]

print("Shape of X_train:", X_train.shape)
print("Shape of X_test:", X_test.shape)

train_indices = set(indices[:n_train])
test_indices = set(indices[n_train:])
print("Union of train and test indices:", train_indices.union(test_indices))

print("Intersection between train and test indices:", train_indices.intersection(test_indices))
```

OUTPUT-

Total Rows: 506

```
Shape of X_train: (455, 13)
```

Shape of X_test: (51, 13)

[illegible]

SPLITTING WITH USAGE OF LIBRARY-

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split

# Load the dataset
data = pd.read_csv("BostonHousing.csv", header='infer')

# Separate features (X) and target (y)
X = data.iloc[:, :-1].values # All columns except the last one
y = data.iloc[:, -1].values  # Last column as target

# Get the total number of rows
nrows = data.shape[0]
print("Total Rows:", nrows)

# Ask for test split ratio
test_split = 0.2

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=test_split)

# Print the shapes of the train and test sets
print("Shapes of the splits:")
print(f"X_train shape: {X_train.shape}, y_train shape: {y_train.shape}")
print(f"X_test shape: {X_test.shape}, y_test shape: {y_test.shape}")
```

Total Rows: 506

Shapes of the splits:

X_train shape: (404, 13), y_train shape: (404,)

X_test shape: (102, 13), y_test shape: (102,)

SPLITTING WITH DIFFERENT TEST SIZE

Outputs-

Size=0.5

```
Total Rows: 506
```

```
Shapes of the splits:
```

```
X_train shape: (253, 13), y_train shape: (253,)
```

```
X_test shape: (253, 13), y_test shape: (253,)
```

Size=0.8

```
Total Rows: 506
```

```
Shapes of the splits:
```

```
X_train shape: (101, 13), y_train shape: (101,)
```

```
X_test shape: (405, 13), y_test shape: (405,)
```