FUNDAMENTALS OF MACHINE LEARNING

LAB ASSIGNMENT - 1

NAME - Kaparotu Venkata Surya Tharani
USN - 22BTRAD018
BRANCH - AI & DE

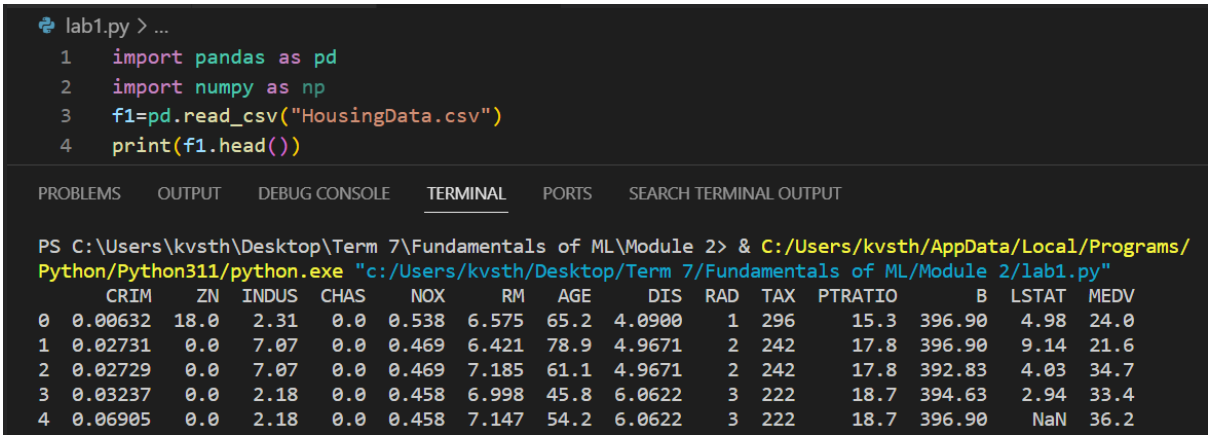
Questions -

1. Load a dataset with missing values (Boston Housing Dataset).

CODE :

```
import numpy as np
import pandas as pd
f1=pd.read_csv(“HousingData.csv”)
print(f1.head())
```

OUTPUT :



```
lab1.py > ...
1  import pandas as pd
2  import numpy as np
3  f1=pd.read_csv("HousingData.csv")
4  print(f1.head())

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS  SEARCH TERMINAL OUTPUT

PS C:\Users\kvsth\Desktop\Term 7\Fundamentals of ML\Module 2> & C:/Users/kvsth/AppData/Local/Programs/
Python/Python311/python.exe "c:/Users/kvsth/Desktop/Term 7/Fundamentals of ML/Module 2/lab1.py"
      CRIM    ZN  INDUS  CHAS    NOX     RM   AGE     DIS  RAD  TAX  PTRATIO       B  LSTAT  MEDV
0  0.00632  18.0   2.31   0.0  0.538  6.575  65.2  4.0900    1  296     15.3  396.90   4.98  24.0
1  0.02731   0.0   7.07   0.0  0.469  6.421  78.9  4.9671    2  242     17.8  396.90   9.14  21.6
2  0.02729   0.0   7.07   0.0  0.469  7.185  61.1  4.9671    2  242     17.8  392.83   4.03  34.7
3  0.03237   0.0   2.18   0.0  0.458  6.998  45.8  6.0622    3  222     18.7  394.63   2.94  33.4
4  0.06905   0.0   2.18   0.0  0.458  7.147  54.2  6.0622    3  222     18.7  396.90    NaN  36.2
```

2. Explore the description of the dataset.

CODE :

```
print (f1.describe)
```

OUTPUT :

```
lab1.py > ...
  6   # Description of the dataset
  7   print(f1.describe)
  8
  9

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS  SEARCH TERMINAL OUTPUT

Python/Python311/python.exe "c:/Users/kvsth/Desktop/Term 7/Fundamentals of ML/Module 2/lab1.py"
<bound method NDFrame.describe of         CRIM    ZN  INDUS  CHAS    NOX     RM   AGE     DIS  RAD  TA
X  PTRATIO       B  LSTAT  MEDV
0    0.00632  18.0   2.31   0.0  0.538  6.575  65.2  4.0900    1  296     15.3  396.90   4.98  24.0
1    0.02731   0.0   7.07   0.0  0.469  6.421  78.9  4.9671    2  242     17.8  396.90   9.14  21.6
2    0.02729   0.0   7.07   0.0  0.469  7.185  61.1  4.9671    2  242     17.8  392.83   4.03  34.7
3    0.03237   0.0   2.18   0.0  0.458  6.998  45.8  6.0622    3  222     18.7  394.63   2.94  33.4
4    0.06905   0.0   2.18   0.0  0.458  7.147  54.2  6.0622    3  222     18.7  396.90    NaN  36.2
..       ...   ...    ...   ...    ...    ...   ...     ...  ...  ...      ...     ...    ...   ...
501  0.06263   0.0  11.93   0.0  0.573  6.593  69.1  2.4786    1  273     21.0  391.99    NaN  22.4
502  0.04527   0.0  11.93   0.0  0.573  6.120  76.7  2.2875    1  273     21.0  396.90   9.08  20.6
503  0.06076   0.0  11.93   0.0  0.573  6.976  91.0  2.1675    1  273     21.0  396.90   5.64  23.9
504  0.10959   0.0  11.93   0.0  0.573  6.794  89.3  2.3889    1  273     21.0  393.45   6.48  22.0
505  0.04741   0.0  11.93   0.0  0.573  6.030   NaN  2.5050    1  273     21.0  396.90   7.88  11.9

[506 rows x 14 columns]>
```

3. Identify the number of missing values corresponding to each feature.

CODE :

```
Missing_values = f1.isnull().sum()
print("Number of missing values for each feature:\n", missing_values)
```

OUTPUT :

```
lab1.py > ...
  9   # Identify missing values
 10   missing_values = f1.isnull().sum()
 11   print("Number of missing values for each feature:\n", missing_values)
 12

PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS  SEARCH TERMINAL OUTPUT

Python/Python311/python.exe "c:/Users/kvsth/Desktop/Term 7/Fundamentals of ML/Module 2/lab1.py"
Number of missing values for each feature:
 CRIM       20
ZN          20
INDUS       20
CHAS        20
NOX          0
RM           0
AGE         20
DIS          0
RAD          0
TAX          0
PTRATIO      0
B            0
LSTAT       20
MEDV         0
dtype: int64
PS C:\Users\kvsth\Desktop\Term 7\Fundamentals of ML\Module 2>
```

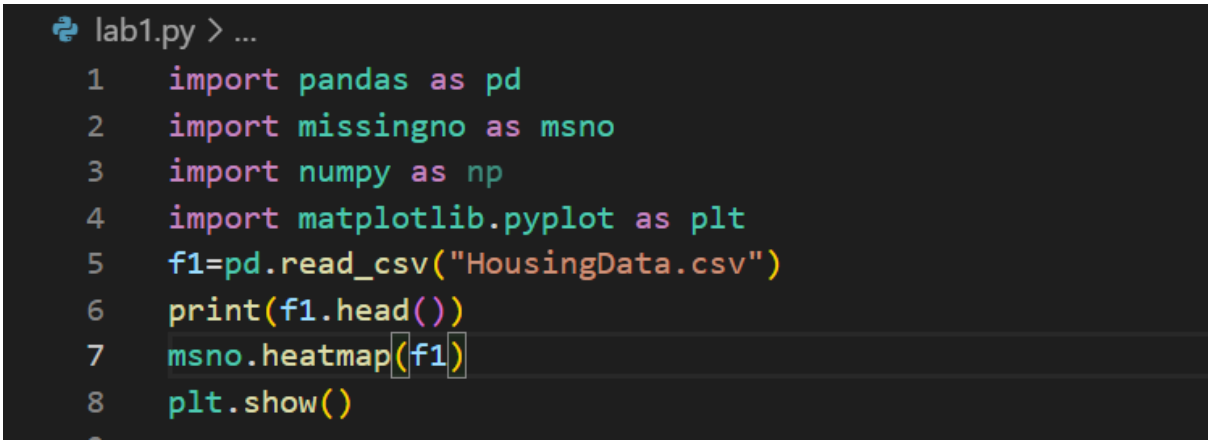
4. Explore and visualize the missing data patterns.

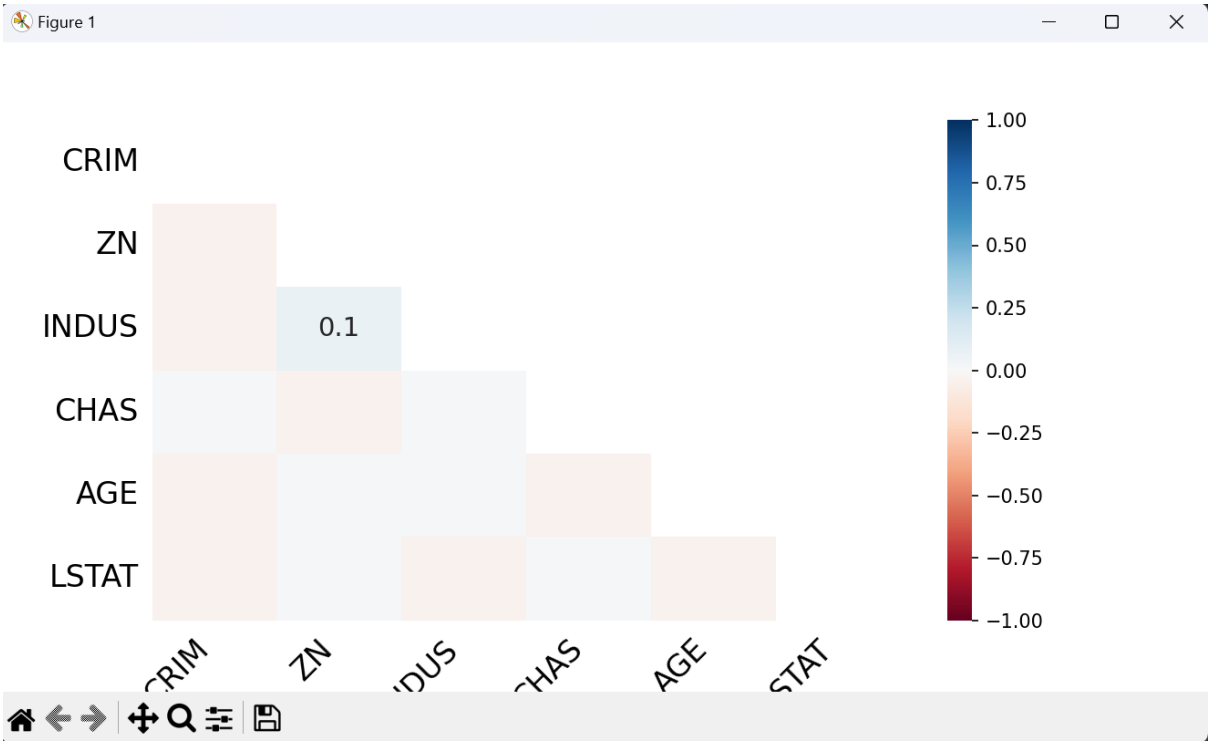
Code :

```
import pandas as pd
import missingno as msno
import numpy as np
import matplotlib.pyplot as plt
f1=pd.read_csv("HousingData.csv")

# visualizing the missing data patterns
msno.heatmap(f1)
plt.show()
```

Output :





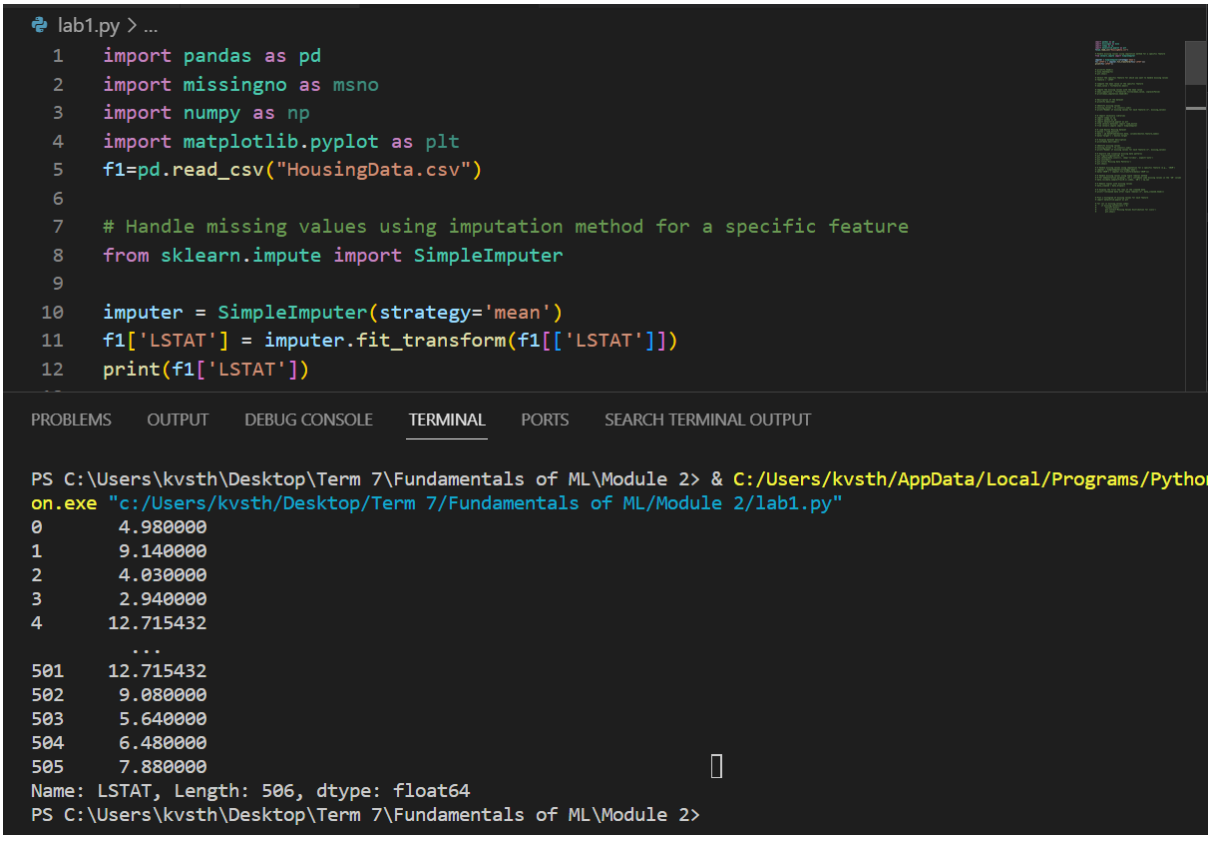
5. Handle missing values using imputation method for a specific feature.

Code :

```
 # Handle missing values using imputation method for a specific feature
from sklearn.impute import SimpleImputer

imputer = SimpleImputer(strategy='mean')
f1['LSTAT'] = imputer.fit_transform(f1[['LSTAT']])
print(f1['LSTAT'])
```

Output :



6. Handle missing values using tuple removal method.

Code :

```
# handling missing values using tuple removal method
f1_dropna = f1.dropna()

# Display the results
print("\nDataset after imputation:\n", f1.head())
print("\nDataset after tuple removal:\n", f1_dropna.head())
```

Output :

lab1.py > ...

```
1   import pandas as pd
2   import missingno as msno
3   import numpy as np
4   import matplotlib.pyplot as plt
5   f1=pd.read_csv("HousingData.csv")
6
7   # handling missing values using tuple removal method
8   f1_dropna = f1.dropna()
9   # Display the results
10  print("\nDataset after imputation:\n", f1.head())
11  print("\nDataset after tuple removal:\n", f1_dropna.head())
12
13
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS SEARCH TERMINAL OUTPUT

```
PS C:\Users\kvsth\Desktop\Term 7\Fundamentals of ML\Module 2> & C:/Users/kvsth/AppData/Local/Programs/P
on.exe "c:/Users/kvsth/Desktop/Term 7/Fundamentals of ML/Module 2/lab1.py"

Dataset after imputation:
      CRIM    ZN  INDUS  CHAS    NOX     RM   AGE     DIS  RAD  TAX  PTRATIO       B  LSTAT  MEDV
0  0.00632  18.0   2.31   0.0  0.538  6.575  65.2  4.0900    1  296     15.3  396.90   4.98  24.0
1  0.02731   0.0   7.07   0.0  0.469  6.421  78.9  4.9671    2  242     17.8  396.90   9.14  21.6
2  0.02729   0.0   7.07   0.0  0.469  7.185  61.1  4.9671    2  242     17.8  392.83   4.03  34.7
3  0.03237   0.0   2.18   0.0  0.458  6.998  45.8  6.0622    3  222     18.7  394.63   2.94  33.4
4  0.06905   0.0   2.18   0.0  0.458  7.147  54.2  6.0622    3  222     18.7  396.90    NaN  36.2

Dataset after tuple removal:
      CRIM    ZN  INDUS  CHAS    NOX     RM   AGE     DIS  RAD  TAX  PTRATIO       B  LSTAT  MEDV
0  0.00632  18.0   2.31   0.0  0.538  6.575  65.2  4.0900    1  296     15.3  396.90   4.98  24.0
1  0.02731   0.0   7.07   0.0  0.469  6.421  78.9  4.9671    2  242     17.8  396.90   9.14  21.6
2  0.02729   0.0   7.07   0.0  0.469  7.185  61.1  4.9671    2  242     17.8  392.83   4.03  34.7
3  0.03237   0.0   2.18   0.0  0.458  6.998  45.8  6.0622    3  222     18.7  394.63   2.94  33.4
5  0.02985   0.0   2.18   0.0  0.458  6.430  58.7  6.0622    3  222     18.7  394.12   5.21  28.7
PS C:\Users\kvsth\Desktop\Term 7\Fundamentals of ML\Module 2>
```