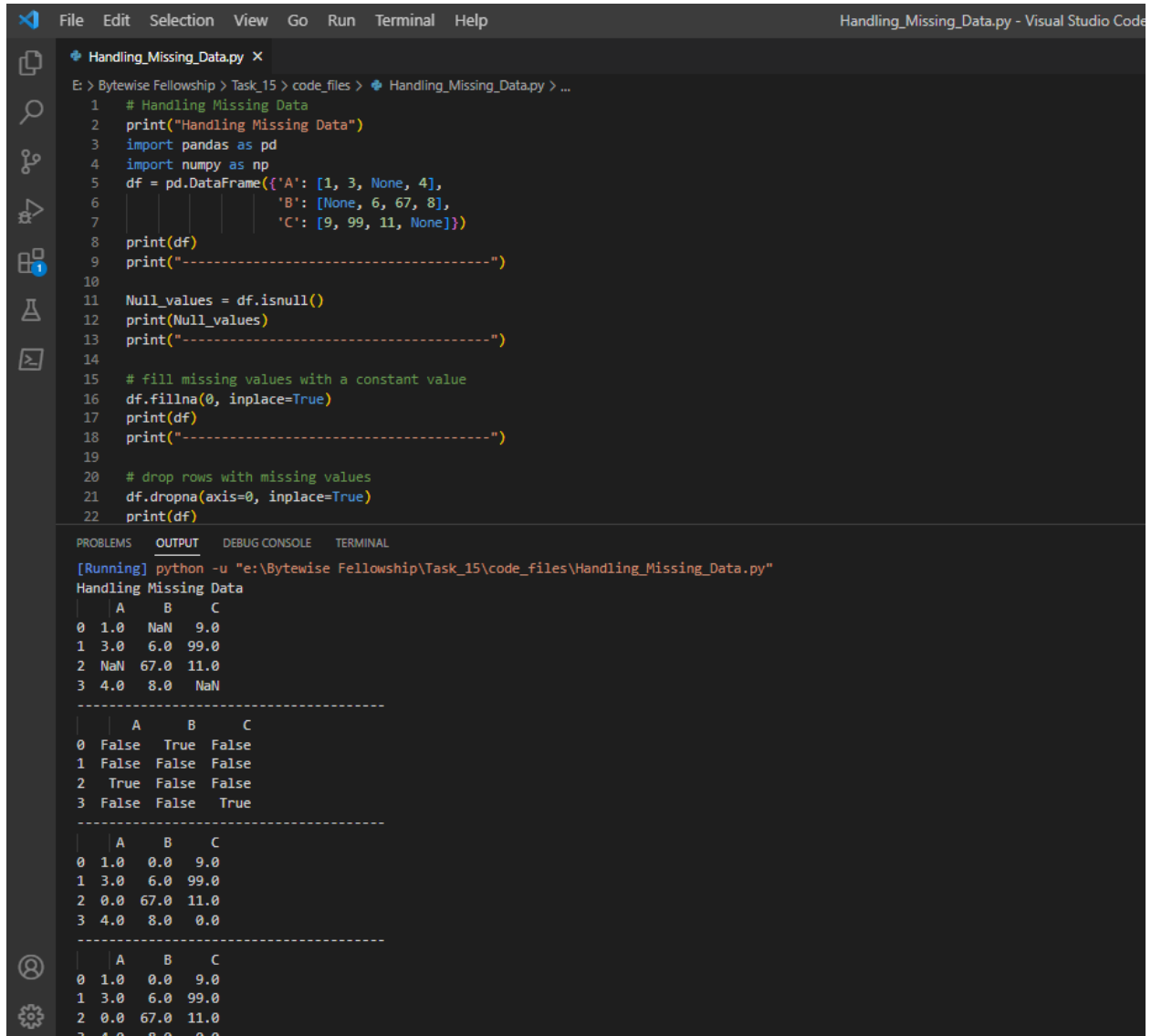


Sania Bibi

Task # 15



```
File Edit Selection View Go Run Terminal Help
Handling_Missing_Data.py - Visual Studio Code

Handling_Missing_Data.py X
E:\> Bytewise Fellowship > Task_15 > code_files > Handling_Missing_Data.py > ...
1 # Handling Missing Data
2 print("Handling Missing Data")
3 import pandas as pd
4 import numpy as np
5 df = pd.DataFrame({'A': [1, 3, None, 4],
6                      'B': [None, 6, 67, 8],
7                      'C': [9, 99, 11, None]})
8 print(df)
9 print("-----")
10
11 Null values = df.isnull()
12 print(Null_values)
13 print("-----")
14
15 # fill missing values with a constant value
16 df.fillna(0, inplace=True)
17 print(df)
18 print("-----")
19
20 # drop rows with missing values
21 df.dropna(axis=0, inplace=True)
22 print(df)
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

[Running] python -u "e:\Bytewise Fellowship\Task_15\code_files\Handling_Missing_Data.py"

Handling Missing Data

	A	B	C
0	1.0	NaN	9.0
1	3.0	6.0	99.0
2	NaN	67.0	11.0
3	4.0	8.0	NaN

	A	B	C
0	False	True	False
1	False	False	False
2	True	False	False
3	False	False	True

	A	B	C
0	1.0	0.0	9.0
1	3.0	6.0	99.0
2	0.0	67.0	11.0
3	4.0	8.0	0.0

	A	B	C
0	1.0	0.0	9.0
1	3.0	6.0	99.0
2	0.0	67.0	11.0
3	4.0	8.0	0.0

File Edit Selection View Go Run Terminal Help Handling_Missing_Data.py - Visual Studio Code

Handling_Missing_Data.py X

E:\> Bytewise Fellowship > Task_15 > code_files > Handling_Missing_Data.py > ...

```
31
32 # Filling and Replacing Values
33
34 print("Filling and Replacing Values")
35
36 # replace specific values
37 df.replace({1: 100, None: 0}, inplace=True)
38 print(df)
39 print("-----")
40
41 # fill missing values using forward fill
42 df.fillna(method='ffill', inplace=True)
43 print(df)
44 print("-----")
45
46 # fill missing values using backward fill
47 df.fillna(method='bfill', inplace=True)
48 print(df)
49 print("=====")
50
51
52 # Removing Duplicates
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

```
2 0.0 67.0 11.0
3 4.0 8.0 0.0
=====
Filling and Replacing Values
|  | A    B    C
0  100.0 0.0  9.0
1   3.0  6.0 99.0
2   0.0 67.0 11.0
3   4.0  8.0  0.0
-----
|  | A    B    C
0  100.0 0.0  9.0
1   3.0  6.0 99.0
2   0.0 67.0 11.0
3   4.0  8.0  0.0
-----
|  | A    B    C
0  100.0 0.0  9.0
1   3.0  6.0 99.0
2   0.0 67.0 11.0
3   4.0  8.0  0.0
=====
Removing Duplicates
```

File Edit Selection View Go Run Terminal HelpHandling_Missing_Data.py - Visual Studio Code

Handling_Missing_Data.py ×

E > Bytewise Fellowship > Task_15 > code_files > Handling_Missing_Data.py > ...

```
48 print(df)
49 print("=====")
50
51
52 # Removing Duplicates
53 print("Removing Duplicates")
54 # dataframe with duplicates
55 df1 = pd.DataFrame({'A': [11, 2, 2, 14],
56                      'B': [5, 6, 6, 13],
57                      'C': [9, 10, 10, 12]})
58 print(df1)
59 print("-----")
60 df1.drop_duplicates(subset=['A', 'B'], inplace=True)
61 df1.drop_duplicates(keep='first', inplace=True)
62 df1.drop_duplicates(keep='last', inplace=True)
63 print(df1)
64 print("=====")
65
66 # Detecting and Removing Outliers
67 print("Detecting and Removing Outliers")
68 df2 = pd.DataFrame({'A': [1, 12, 13, 4, 5, 100],
69                      'B': [5, 6, 7, 10, 0, 100]})
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

	A	B	C
0	100.0	0.0	9.0
1	3.0	6.0	99.0
2	0.0	67.0	11.0
3	4.0	8.0	0.0

=====

Removing Duplicates

	A	B	C
0	11	5	9
1	2	6	10
2	2	6	10
3	14	13	12

	A	B	C
0	11	5	9
1	2	6	10
3	14	13	12

=====

Detecting and Removing Outliers

	A	B	C
0	1	5	9
1	12	6	10
2	13	7	11
3	4	18	12

FileEditSelectionViewGoRunTerminalHelp

Handling_Missing_Data.py - Visual Studio Code

Handling_Missing_Data.py x

E:\> ByteWise Fellowship > Task_15 > code_files > Handling_Missing_Data.py > ...

```
65
66 # Detecting and Removing Outliers
67 print("Detecting and Removing Outliers")
68 df2 = pd.DataFrame({'A': [1, 12, 13, 4, 5, 100],
69                    'B': [5, 6, 7, 10, 9, 100],
70                    'C': [9, 10, 11, 12, 13, 100]})
71 print(df2)
72 print("-----")
73
74 # Z-score for each data point
75 z_scores = np.abs((df2 - df2.mean()) / df2.std())
76 print(z_scores)
77 print("-----")
78 threshold = 3
79 outliers = z_scores > threshold
80
81 # remove outliers from the data
82 df2 = df2[~outliers.any(axis=1)]
83 print(df2)
84
85
--
```

PROBLEMS

OUTPUT

DEBUG CONSOLE

TERMINAL

Detecting and Removing Outliers

	A	B	C
0	1	5	9
1	12	6	10
2	13	7	11
3	4	10	12
4	5	9	13
5	100	100	100

	A	B	C
0	0.562007	0.511855	0.462942
1	0.274469	0.485149	0.435441
2	0.248329	0.458444	0.407939
3	0.483588	0.164684	0.380438
4	0.457448	0.405033	0.352936
5	2.025841	2.025165	2.039697

	A	B	C
0	1	5	9
1	12	6	10
2	13	7	11
3	4	10	12
4	5	9	13
5	100	100	100