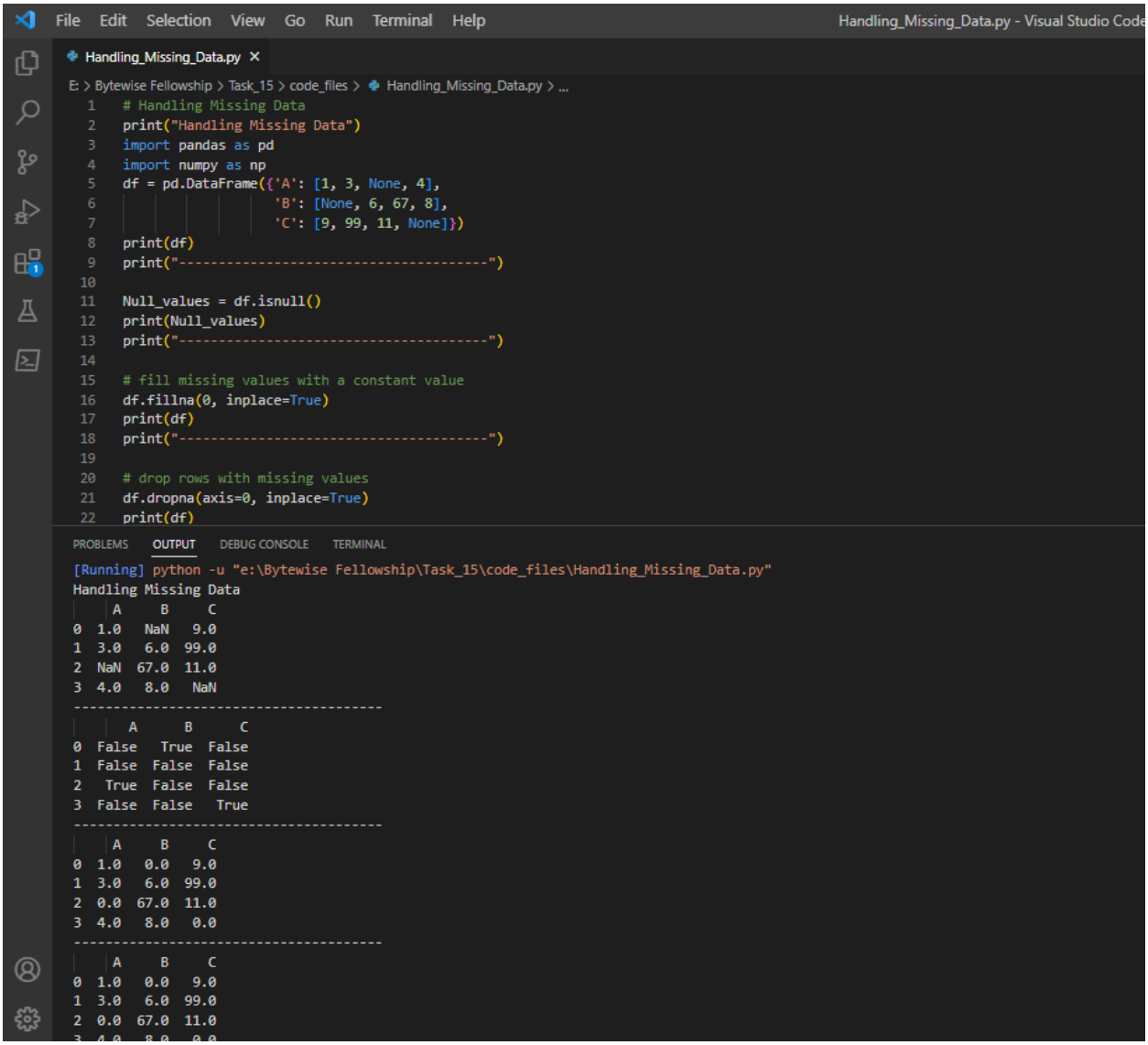
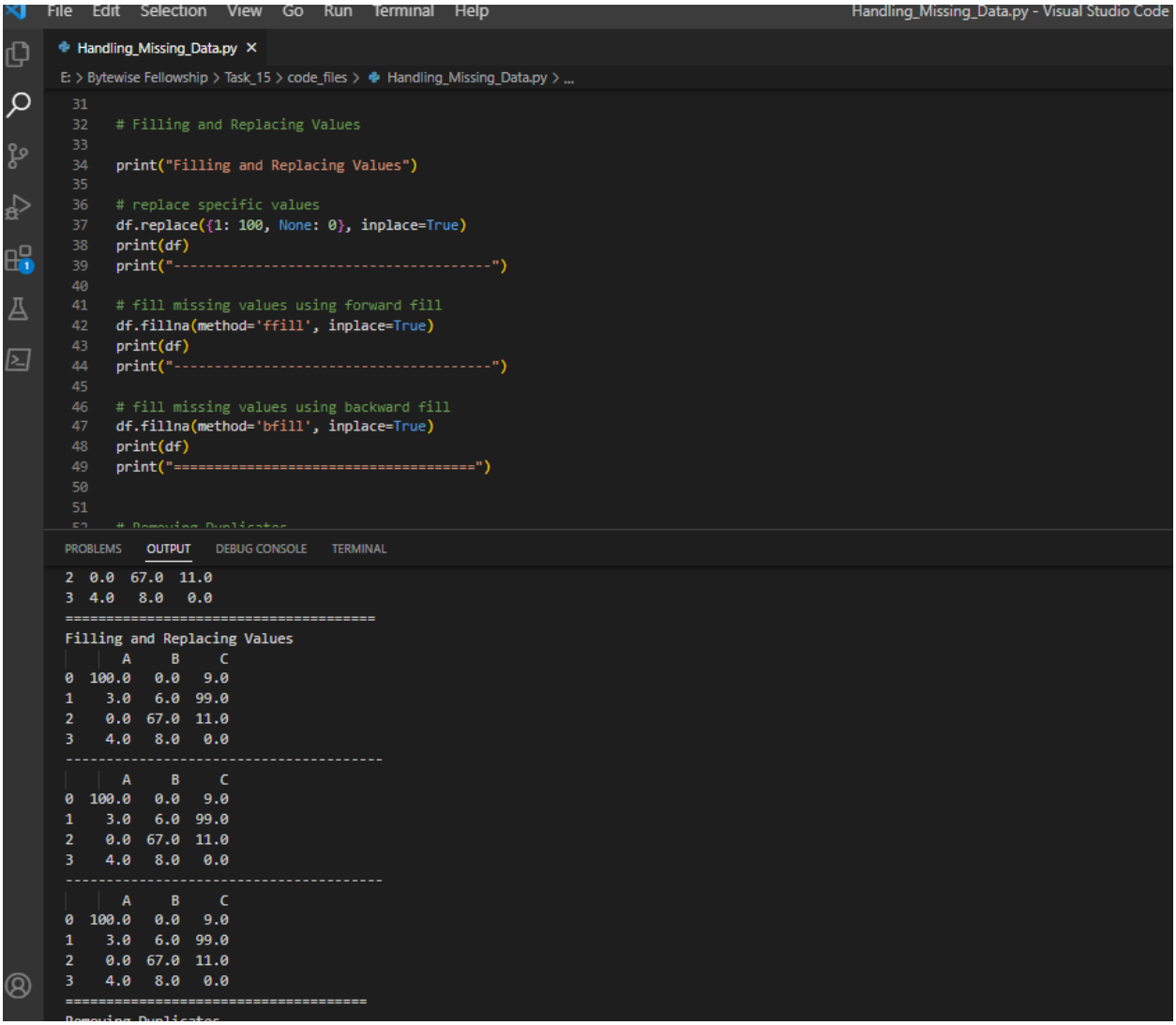
Sania Bibi

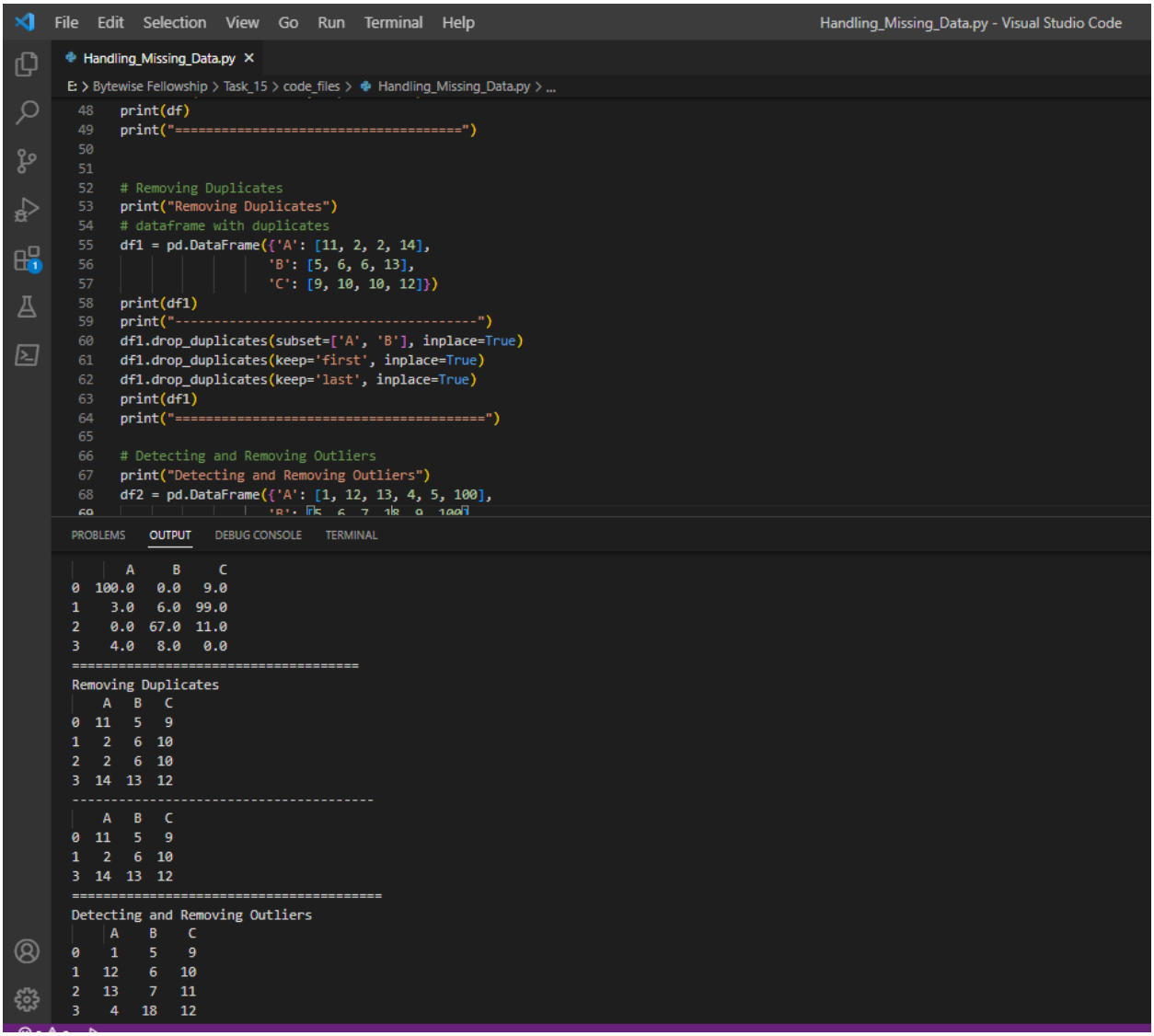
Task # 15



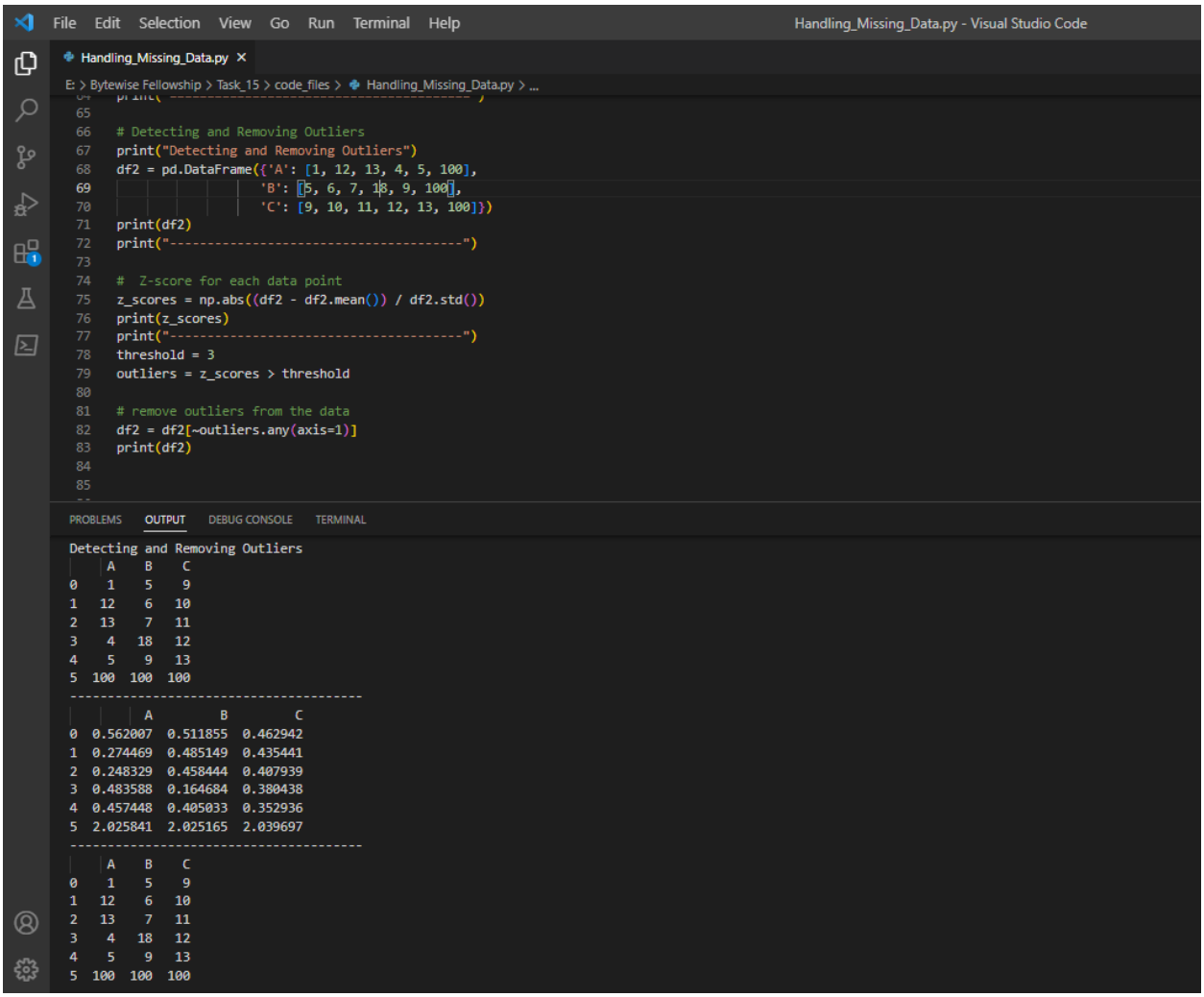

File Edit Selection View Go Run Terminal Help
Handling_Missing_Data.py - Visual Studio Code
Handling_Missing_Data.py
E: > Bytewise Fellowship > Task_15 > code_files > Handling_Missing_Data.py > ...
31
32 # Filling and Replacing Values
33
34 print("Filling and Replacing Values")
35
36 # replace specific 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
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
==========


File Edit Selection View Go Run Terminal Help
Handling_Missing_Data.py - Visual Studio Code
Handling_Missing_Data.py
E: > Bytewise Fellowship > Task_15 > code_files > Handling_Missing_Data.py > ...
print(df)
print("====================================")
# Removing Duplicates
print("Removing Duplicates")
# dataframe with duplicates
df1 = pd.DataFrame({'A': [11, 2, 2, 14],
'B': [5, 6, 6, 13],
'C': [9, 10, 10, 12]})
print(df1)
print("------------------------------------")
df1.drop_duplicates(subset=['A', 'B'], inplace=True)
df1.drop_duplicates(keep='first', inplace=True)
df1.drop_duplicates(keep='last', inplace=True)
print(df1)
print("====================================")
# Detecting and Removing Outliers
print("Detecting and Removing Outliers")
df2 = pd.DataFrame({'A': [1, 12, 13, 4, 5, 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


File Edit Selection View Go Run Terminal Help
Handling_Missing_Data.py - Visual Studio Code
Handling_Missing_Data.py
E: > Bytewise Fellowship > Task_15 > code_files > Handling_Missing_Data.py > ...
# Detecting and Removing Outliers
print("Detecting and Removing Outliers")
df2 = pd.DataFrame({'A': [1, 12, 13, 4, 5, 100],
'B': [5, 6, 7, 18, 9, 100],
'C': [9, 10, 11, 12, 13, 100]})
print(df2)
print("----------")
# Z-score for each data point
z_scores = np.abs((df2 - df2.mean()) / df2.std())
print(z_scores)
print("----------")
threshold = 3
outliers = z_scores > threshold
# remove outliers from the data
df2 = df2[~outliers.any(axis=1)]
print(df2)
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 18 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 18 12
4 5 9 13
5 100 100 100