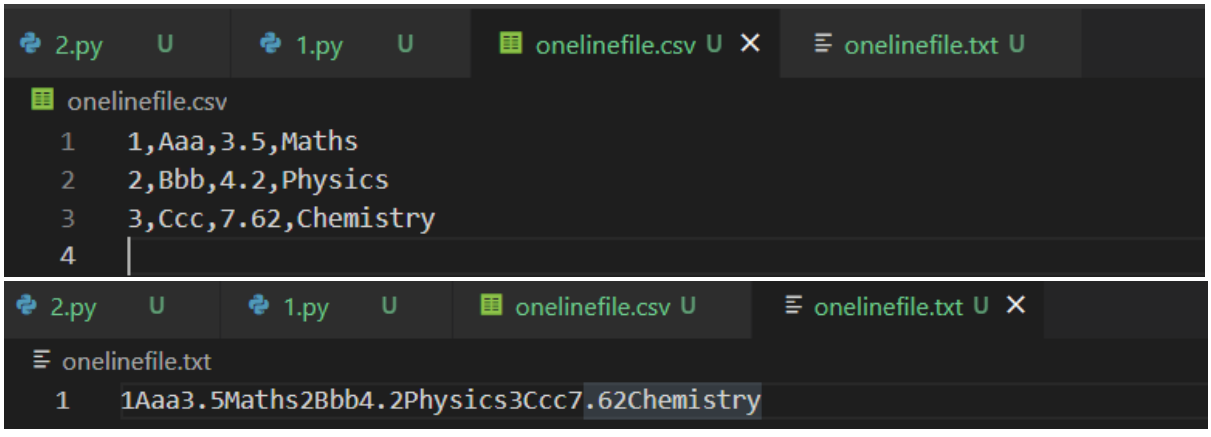
Sree Chandana.M **CH.EN.U4CSE20042**

1)

```
'''1. Write a python program that reads the contents from the given file
"onelinefile.txt". The file contains a single line which is of the format
(int)(string)(float)(string) repeatedly. For e.g.
1Aaa3.5Maths2Bbb4.2Physics3Ccc7.62Chemistry
Your main task is to split the contents of the given file based on their
format and write it into a .csv file say "Filename2.csv". For e.g. the above
txt file should be converted into a csv file such that the contents look like
this:
1,Aaa,3.5,Maths
2,Bbb,4.2,Physics
3,Ccc,7.62,Chemistry
Contents of "onelinefile.txt"
1Aaa3.5Maths2Bbb4.2Physics3Ccc7.62Chemistry4Ddd9.55Biology5Eee4.0Social6Fff7.6
English7Ggg3.111Maths8Hhh9.99Physics9Iii1.23Civics'''
import re,csv
f = open('onelinefile.txt')
for i in f:
        a = re.findall(r'[+-]?[0-9]+\.[0-9]+', i)
        b = re.findall(r'[a-zA-Z]+', i)
        c = 0
        for p in range(len(a)):
            with open('onelinefile.csv', 'a', newline='') as file:
                writer = csv.writer(file)
                writer.writerow([str(p+1), b[c],a[p],b[c+1]])
            c += 2
with open('onelinefile.csv', 'r',) as file:
    reader = csv.reader(file)
    for row in reader:
        print(','.join(row))
```

Files:



Output:

```
1,Aaa,3.5,Maths
2,Bbb,4.2,Physics
3,Ccc,7.62,Chemistry
```

2)

```
'''2.Data formatting
Python libraries represent missing numbers as nan which is short for “not a
number”. Most libraries (including scikit-learn) will give you an error if you
try to build a model using data with missing values. One of the common
solution to get around this issue is to impute or fill in the missing value
with a number or value
of same format. From the given dataset, find the missing values(Nan/NA/-/Nil)
and change those values into an appropriate number.'''
```

File:

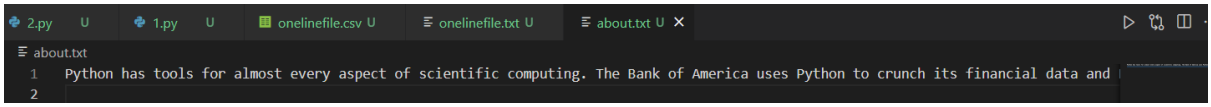
Output:

3)

```
'''3.Read the file 'about.txt' and find the words with atleast 6 letters and
the most frequently used word.
Contents of the file 'about.txt':
Python has tools for almost every aspect of scientific computing. The Bank of
America uses Python to crunch its financial data and Facebook looks upon the
Python library Pandas for its data analysis.
While there are many libraries available to perform data analysis in Python,
here are a few: NumPy, SciPy, Pandas and Matplotlib.'''
count = 0
word = ""
max_count = 0
words = []
file = open("about.txt","r")
for line in file:
    string = line.lower().replace(',','').replace('.','').split(" ")
    for s in string:
        if len(s) == 6:
            words.append(s);
for i in range(0, len(words)):
    count = 1
    for j in range(i+1, len(words)):
        if(words[i] == words[j]):
            count = count + 1
    if(count > max_count):
        max_count = count
        word = words[i]
```

```
print("Most repeated word:" + word)
file.close()
```

File:



Output:

```
Most repeated word:python
```