



COGNIZANCE 2022

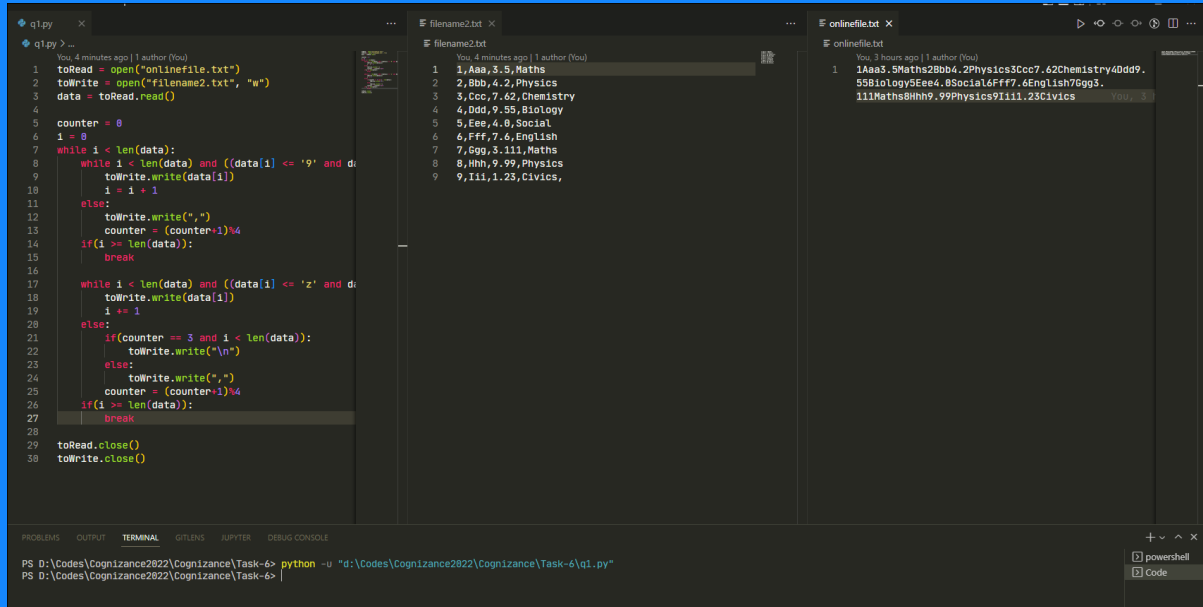
Task-6



JUNE 30, 2022
SAYAN BANERJEE
CH.EN.U4CSE20064

DATA HANDLING

Q1



```
1 toRead = open("onlinefile.txt")
2 toWrite = open("filename2.txt", "w")
3 data = toRead.read()
4
5 counter = 0
6 i = 0
7 while i < len(data):
8     while i < len(data) and ((data[i] <= '9' and data[i] >= '0') or data[i] == ','):
9         toWrite.write(data[i])
10        i = i + 1
11    else:
12        toWrite.write(",")
13        counter = (counter+1)%4
14    if(i >= len(data)):
15        break
16
17    while i < len(data) and ((data[i] <= 'z' and data[i] >= 'a') or data[i] == ','):
18        toWrite.write(data[i])
19        i = i + 1
20    else:
21        if(counter == 3 and i < len(data)):
22            toWrite.write("\n")
23        else:
24            toWrite.write(",")
25            counter = (counter+1)%4
26    if(i >= len(data)):
27        break
28
29 toRead.close()
30 toWrite.close()
```

filename2.txt

```
1 1,Aaa,3.5,Maths
2 2,Bbb,4.2,Physics
3 3,Ccc,7.62,Chemistry
4 4,Ddd,9.55,Biology
5 5,Eee,4.8,Social
6 6,Fff,7.6,English
7 7,Ggg,3.111,Maths
8 8,Hhh,9.99,Physics
9 9,Iii,1.23,Civics
```

onlinefile.txt

```
1 1Aaa3.5Maths2Bbb4.2Physics3Ccc7.62Chemistry4Ddd9.55Biology5Eee4.8Social6Fff7.6English7Ggg3.111Maths8Hhh9.99Physics9Iiii.23Civics
```

Q2

Reading the data set into a list of dictionaries

```
In [ ]: from csv import DictReader
import pandas as pd

with open("dataset.csv") as file:
    data = list(DictReader(file))
```

Finding out all the columns that have the potential nil values.

```
In [ ]: list = []
for k in data[0].keys():
    for df in data:
        if(df[k] == "NA" or df[k] == "nil" or df[k] == "" or df[k] == "nan"):
            list.append(k)
            break
print(list)

['LotFrontage', 'Alley', 'BsmtQual', 'BsmtCond', 'BsmtExposure', 'BsmtFinType1', 'BsmtFinType2']
```

Replacing all the nil or nan values with 0

```
In [ ]: for df in data:
    for l in list:
        if(df[l] == "NA" or df[l] == "nil" or df[l] == "" or df[l] == "nan"):
            df[l] = "0"
```

Testing if the data is nil free

```
In [ ]: list = []
for k in data[0].keys():
    for df in data:
        if(df[k] == "NA" or df[k] == "nil" or df[k] == "" or df[k] == "nan"):
            list.append(k)
            break
print(list)

[]
```

Q3.

PROBLEMS OUTPUT TERMINAL GITLENS JUPYTER DEBUG CONSOLE

```
PS D:\Codes\Cognizance2022\Cognizance\Task-6> python -u "d:\Codes\Cognizance2022\Cognizance\Task-6\q3.py"
```

```
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.
```

```
The most used word is: Python used 3 times.
```

```
PS D:\Codes\Cognizance2022\Cognizance\Task-6> |
```