CHALLA SHIVA SHANKAR KOUSHIK REDDY

CH.EN.U4CSE20015

DISCORD HANDLE: KOUSHIK#4875

TASK 6 [PYTHON - MEDICORE LVL]

QUESTION – 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 eg

```
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
```

OUTPUT

QUESTION – 2

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 solutions 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.

OUTPUT

```python
import pandas as pd
import numpy as np
```

[1] Python

```python
df = pd.read_csv("https://raw.githubusercontent.com/cognizance-amrita/AI-Tasks/main/Task-1/Q2-Dataset.csv")
```

[2] Python

```python
df.head()
```

[3] Python

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	MasVnrArea	ExterQual	ExterCond	Foundation	Bsmt
0	1	60	RL	65.0	8450	Pave	NaN	Reg	Lvl	AllPub	...	196	Gd	TA	PConc	
1	2	20	RL	80.0	9600	Pave	NaN	Reg	Lvl	AllPub	...	0	TA	TA	CBlock	
2	3	60	RL	68.0	11250	Pave	NaN	IR1	Lvl	AllPub	...	162	Gd	TA	PConc	
3	4	70	RL	60.0	9550	Pave	NaN	IR1	Lvl	AllPub	...	0	TA	TA	BrkTil	
4	5	60	RL	84.0	14260	Pave	NaN	IR1	Lvl	AllPub	...	350	Gd	TA	PConc	

5 rows × 36 columns

```python
missing_value_formats = ["n.a.","?","NA","n/a", "na", "--"]
df = pd.read_csv("https://raw.githubusercontent.com/cognizance-amrita/AI-Tasks/main/Task-1/Q2-Dataset.csv", na_values = missing_value_formats)
```

[32] Python

```python
print(df.isnull().sum())
```

[6] Python

```
Output exceeds the size limit. Open the full output data in a text editor
Id               0
MSSubClass       0
MSZoning         0
LotFrontage     14
LotArea          0
Street           0
Alley           93
LotShape         0
LandContour      0
Utilities        0
LotConfig        0
LandSlope        0
Neighborhood     0
Condition1       0
BsmtFinSF1       0
BsmtFinType2     3
dtype: int64
```

```python
df['LotFrontage'].fillna(1, inplace=True)
```

[7] Python

```python
print(df['LotFrontage'])
```

[8] Python

```
0     65.0
1     80.0
2     68.0
3     60.0
4     84.0
      ...
94    69.0
95     1.0
96    78.0
97    73.0
98    85.0
Name: LotFrontage, Length: 99, dtype: float64
```

```
print(df['Alley'].isnull())
```

[9] Python

```
0     True
1     True
2     True
3     True
4     True
      ...
94    True
95    True
96    True
97    True
98    True
Name: Alley, Length: 99, dtype: bool
```

```
df['Alley'].fillna('no alley name mentioned', inplace=True)
print(df['Alley'])
```

[10] Python

```
0     no alley name mentioned
1     no alley name mentioned
2     no alley name mentioned
94    no alley name mentioned
95    no alley name mentioned
96    no alley name mentioned
97    no alley name mentioned
98    no alley name mentioned
Name: Alley, Length: 99, dtype: object
```

```
print(df['BsmtQual'].isnull())
```

[11] Python

```
0     False
1     False
2     False
3     False
4     False
      ...
94    False
95    False
96    False
97    False
98    False
Name: BsmtQual, Length: 99, dtype: bool
```

```
df[df['BsmtQual'].isnull()]
```

[12] Python

Run and Debug (Ctrl+Shift+D)

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	MasVnrArea	ExterQual	ExterCond	Foundation
17	18	90	RL	72.0	10791	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	Slab
39	40	90	RL	65.0	6040	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	PConc
90	91	20	RL	60.0	7200	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	Slab

3 rows × 36 columns

```
df['BsmtQual'].fillna('3', inplace=True)
```

[33] Python

```
df[df['BsmtQual'].isnull()]
```

[14] Python

```
[14] df[df['BsmtQual'].isnull()]
```

Python

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	MasVnrArea	ExterQual	ExterCond	Foundation	BsmtC

0 rows × 36 columns

```
[15] print(df['BsmtCond'].isnull())
```

Python

```
0     False
1     False
2     False
3     False
4     False
      ...
94    False
95    False
96    False
97    False
98    False
```

```
[16] df[df['BsmtCond'].isnull()]
```

Python

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	MasVnrArea	ExterQual	ExterCond	Foundation
17	18	90	RL	72.0	10791	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	Slab
39	40	90	RL	65.0	6040	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	PConc
90	91	20	RL	60.0	7200	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	Slab

3 rows × 36 columns

```
[34] df['BsmtCond'].fillna('78', inplace=True)
```

Python

```
[18] df[df['BsmtCond'].isnull()]
```

Python

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	MasVnrArea	ExterQual	ExterCond	Foundation	BsmtC

```
[20] df[df['BsmtExposure'].isnull()]
```

Python

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	MasVnrArea	ExterQual	ExterCond	Foundation
17	18	90	RL	72.0	10791	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	Slab
39	40	90	RL	65.0	6040	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	PConc
90	91	20	RL	60.0	7200	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	Slab

3 rows × 36 columns

```
[35] df['BsmtExposure'].fillna('no exposure mentioned', inplace=True)
```

Python

```
df[df['BsmtExposure'].isnull()]
```

[35]

```
df[df['BsmtExposure'].isnull()]
```

[22]

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	MasVnrArea	ExterQual	ExterCond	Foundation	BsmtC

0 rows × 36 columns

```
print(df['BsmtFinType1'].isnull())
```

[23]

```
0     False
1     False
2     False
3     False
4     False
      ...
94    False
95    False
96    False
97    False
98    False
```

```
df[df['BsmtFinType1'].isnull()]
```

[24]

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	MasVnrArea	ExterQual	ExterCond	Foundation
17	18	90	RL	72.0	10791	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	Slab
39	40	90	RL	65.0	6040	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	PConc
90	91	20	RL	60.0	7200	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	Slab

3 rows × 36 columns

```
df['BsmtFinType1'].fillna('Values not mentioned', inplace=True)
```

[36]

```
df[df['BsmtFinType1'].isnull()]
```

[26]

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	MasVnrArea	ExterQual	ExterCond	Foundation	BsmtC

0 rows × 36 columns

```
print(df['BsmtFinType2'].isnull())
```

[27]

```
0     False
1     False
2     False
3     False
4     False
      ...
94    False
95    False
96    False
97    False
98    False
Name: BsmtFinType2, Length: 99, dtype: bool
```

```
df[df['BsmtFinType2'].isnull()]
```

```
df[df['BsmtFinType2'].isnull()]
```

[28] Python

	Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	MasVnrArea	ExterQual	ExterCond	Foundation
17	18	90	RL	72.0	10791	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	Slab
39	40	90	RL	65.0	6040	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	PConc
90	91	20	RL	60.0	7200	Pave	no alley name mentioned	Reg	Lvl	AllPub	...	0	TA	TA	Slab

3 rows × 36 columns

```
df['BsmtFinType2'].fillna('57', inplace=True)
```

[37] Python

```
df[df['BsmtFinType2'].isnull()]
```

```
df[df['BsmtFinType2'].isnull()]
```

[30] Python

Id	MSSubClass	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	...	MasVnrArea	ExterQual	ExterCond	Foundation	BsmtC

0 rows × 36 columns

+ Code + Markdown

```
print(df.isnull().sum())
```

[31] Python

```
Output exceeds the size limit. Open the full output data in a text editor
Id             0
MSSubClass     0
MSZoning       0
LotFrontage    0
LotArea        0
Street         0
Alley          0
LotShape       0
LandContour    0
Utilities      0
LotConfig      0
LandSlope      0
```

```
LandSlope       0
Neighborhood    0
Condition1      0
Condition2      0
BldgType        0
HouseStyle      0
OverallQual     0
OverallCond     0
YearBuilt       0
YearRemodAdd    0
RoofStyle       0
RoofMatl        0
Exterior1st     0
Exterior2nd     0
...
BsmtFinType1    0
BsmtFinSF1      0
BsmtFinType2    0
dtype: int64
```

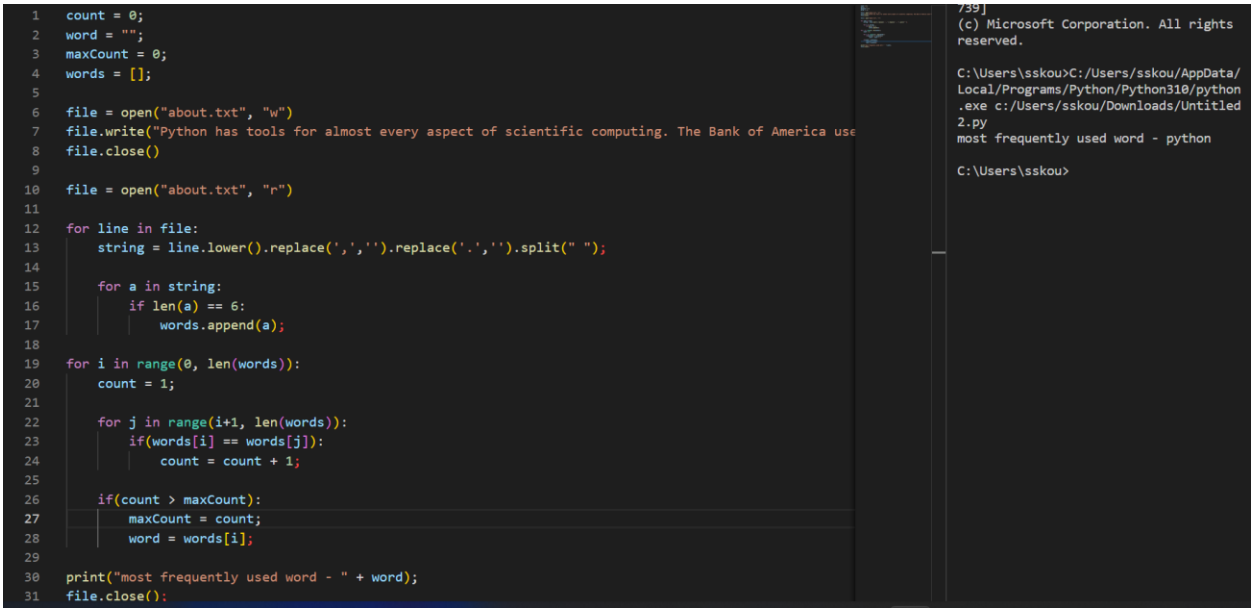
QUESTION – 3

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.”

OUTPUT



```
count = 0;
word = "";
maxCount = 0;
words = [];

file = open("about.txt", "w")
file.write("Python has tools for almost every aspect of scientific computing. The Bank of America use
file.close()

file = open("about.txt", "r")

for line in file:
    string = line.lower().replace(',','').replace('.','').split(" ");

    for a in string:
        if len(a) == 6:
            words.append(a);

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 > maxCount):
        maxCount = count;
        word = words[i];

print("most frequently used word - " + word);
file.close();
```

```
739]
(c) Microsoft Corporation. All rights
reserved.

C:\Users\sskou>C:/Users/sskou/AppData/
Local/Programs/Python/Python310/python
.exe c:/Users/sskou/Downloads/Untitled
2.py
most frequently used word - python

C:\Users\sskou>
```