ASHUTOSH P

CH.EN.U4CSE20007

COGNIZANCE TASK - 6.

Question - 1: Write a python program that reads the contents from the given file 'onelinefile.txt'.Split the contents of the given file based on their format and write it into a .csv file say 'Filename2.csv'.

File name: Task6q1.py

Output:

```
=================== RESTART: C:\Amrita\Cognizance\Task6q1.py ===================
Contents of the file:
1,Aaa,3.5,Maths
2,Bbb,4.2,Physics
3,Ccc,7.62,Chemistry
4,Ddd,9.55,Biology
5,Eee,4.0,Social
6,Fff,7.6,English
7,Ggg,3.111,Maths
8,Hhh,9.99,Physics
9,Iii,1.23,Civics
```

Question - 2: From the given dataset, find the missing values(Nan/NA/-/Nil) and change those values into an appropriate number.

File name: Task6q2.py

Output:

```
================== RESTART: C:\Amrita\Cognizance\Task6q2.py ==================

Missing values in the csv file:
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
Condition2         0
BldgType           0
HouseStyle         0
OverallQual        0
OverallCond        0
YearBuilt          0
YearRemodAdd       0
RoofStyle          0
RoofMatl           0
Exterior1st        0
Exterior2nd        0
MasVnrType         0
MasVnrArea         0
ExterQual          0
ExterCond          0
Foundation         0
BsmtQual           3
BsmtCond           3
BsmtExposure       3
BsmtFinType1       3
BsmtFinSF1         0
BsmtFinType2       3
dtype: int64

Missing values in LotFrontage:

0      False
1      False
2      False
3      False
```

```
2       False
3       False
4       False
        ...
94      False
95       True
96      False
97      False
98      False
Name: LotFrontage, Length: 99, dtype: bool

updated LotFrontage values(changed to '-1' for LotFrontage instead of NA):

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

Missing values in Alley:

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

updated Alley values(changed to 'empty' for Alley instead of NA):

0       empty_Alley
1       empty_Alley
2       empty_Alley
3       empty_Alley
```

```
2      empty_Alley
3      empty_Alley
4      empty_Alley
          ...
94     empty_Alley
95     empty_Alley
96     empty_Alley
97     empty_Alley
98     empty_Alley
Name: Alley, Length: 99, dtype: object

 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

updated BsmtQual values(changed to 'empty' for BsmtQual instead of NA):

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

 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

updated BsmtCond values(changed to 'empty' for BsmtCond instead of NA):

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

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

updated BsmtExposure values(changed to 'empty' for BsmtExposure instead of NA):

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

```
 0      False
1       False
2       False
3       False
4       False
        ...
94      False
95      False
96      False
97      False
98      False
Name: BsmtFinType1, Length: 99, dtype: bool

updated BsmtFinType1 values(changed to 'empty' for BsmtFinType1 instead of NA):

0       False
1       False
2       False
3       False
4       False
        ...
94      False
95      False
96      False
97      False
98      False
Name: BsmtFinType1, Length: 99, dtype: bool

 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

updated BsmtFinType2 values(changed to 'empty' for BsmtFinType2 instead of NA):

0       False
1       False
2       False
3       False
```

```
95     False
96     False
97     False
98     False
Name: BsmtFinType2, Length: 99, dtype: bool


Updated csv file:

Id                 0
MSSubClass         0
MSZoning           0
LotFrontage        0
LotArea            0
Street             0
Alley              0
LotShape           0
LandContour        0
Utilities          0
LotConfig          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
MasVnrType         0
MasVnrArea         0
ExterQual          0
ExterCond          0
Foundation         0
BsmtQual           0
BsmtCond           0
BsmtExposure       0
BsmtFinType1       0
BsmtFinSF1         0
BsmtFinType2       0
dtype: int64
```

Question - 3: Read the file 'about.txt' and find the words with at least 6 letters and the most frequently used word.

File name: Task6q3.py

Output:

```
=================== RESTART: C:\Amrita\Cognizance\Task6q3.py ===================
Python
almost
aspect
scientific
computing
America
Python
crunch
financial
Facebook
Python
library
Pandas
analysis
libraries
available
perform
analysis
Python
Pandas
Matplotlib

Most repeated word over 6 letters long: python
```