Havirbhavi. P

Task-6:

```
Question-1

File Handling is one of the basic important task when it comes to building machine learning models or neural networks. Building a good model always starts with finding datasets and processing it, for which, file handling acts as a stepping stone.
```

Notepad contents: (onefileline.txt)

```
1Aaa3.5Maths2Bbb4.2Physics3Ccc7.62Chemistry4Ddd9.55Biology5Eee4.0Social6Fff7.6English7Ggg3.111Maths8Hhh9.99Physics9Iii1.23Civics
```

OUTPUT:

```
============= RESTART: C:/Users/P Havirbhavi/OneDrive/Desktop/1.py =============
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

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.
```

Output:

```
========== RESTART: C:/Users/P Havirbhavi/OneDrive/Desktop/Task6-2.p
0     False
1     False
2     False
3     False
4     False
      ...
94    False
95     True
96    False
97    False
98    False
Name: LotFrontage, Length: 99, dtype: bool
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
```

```
BsmtFinSF1       0
BsmtFinType2     3
dtype: int64
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
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
0     no alley here
1     no alley here
2     no alley here
3     no alley here
4     no alley here
          ...
94    no alley here
95    no alley here
96    no alley here
97    no alley here
98    no alley here
Name: Alley, Length: 99, dtype: object
0     False
1     False
2     False
3     False
4     False
      ...
94    False
95    False
```

```
3     no alley here
4     no alley here
          ...
94    no alley here
95    no alley here
96    no alley here
97    no alley here
98    no alley here
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
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
0     False
1     False
```

```
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
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
```

```
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 atleast 6 letters and the most frequently used word.

Notepad Contents : (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:

```
============= RESTART: C:/Users/P Havirbhavi/OneDrive/Desktop/3.py =============
Most frequently used word is : python
```