**Q1**

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.

**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

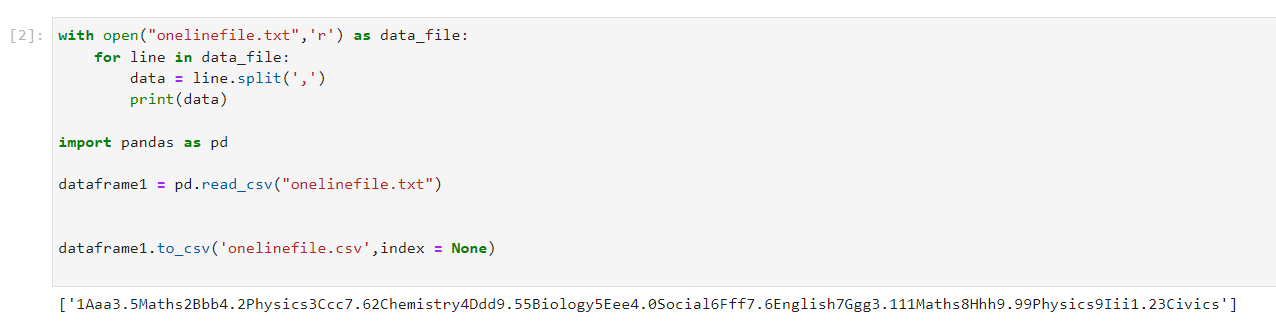
**Your main task is to split the contents of the given file based on their format and write it into a .csv file say 'Filename2.csv'. For e.g. the above txt file should be converted into a csv file such that the contents look like this:**

1,Aaa,3.5,Maths

2,Bbb,4.2,Physics

3,Ccc,7.62,Chemistry

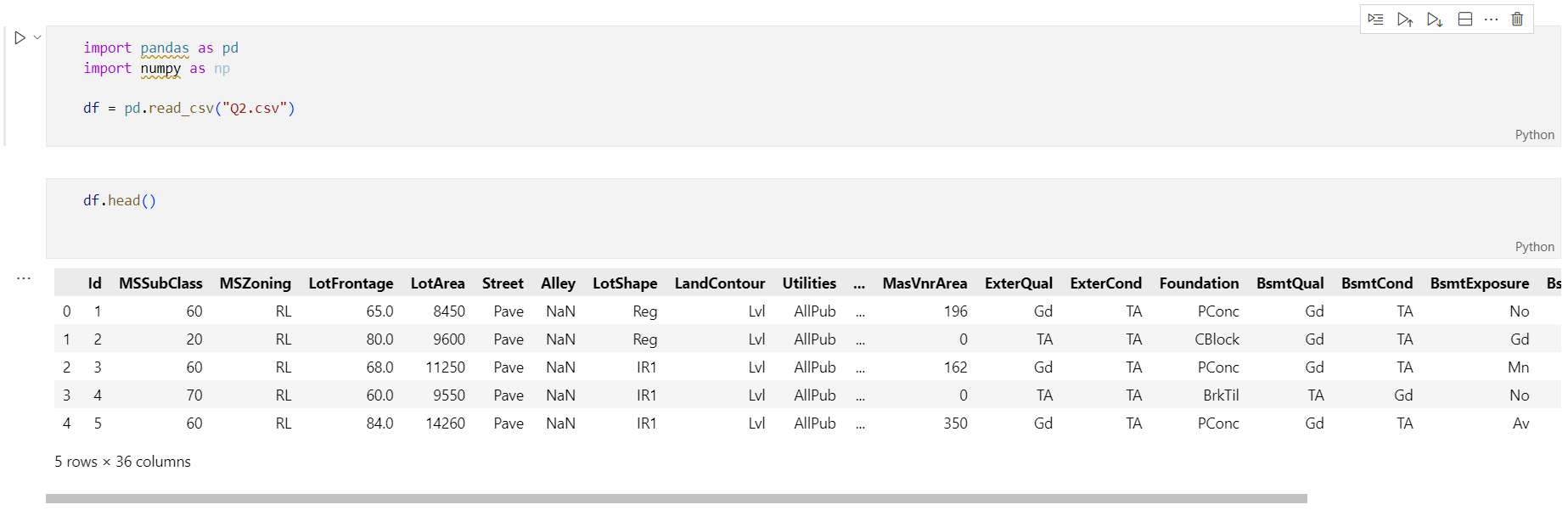
**SOLN:**

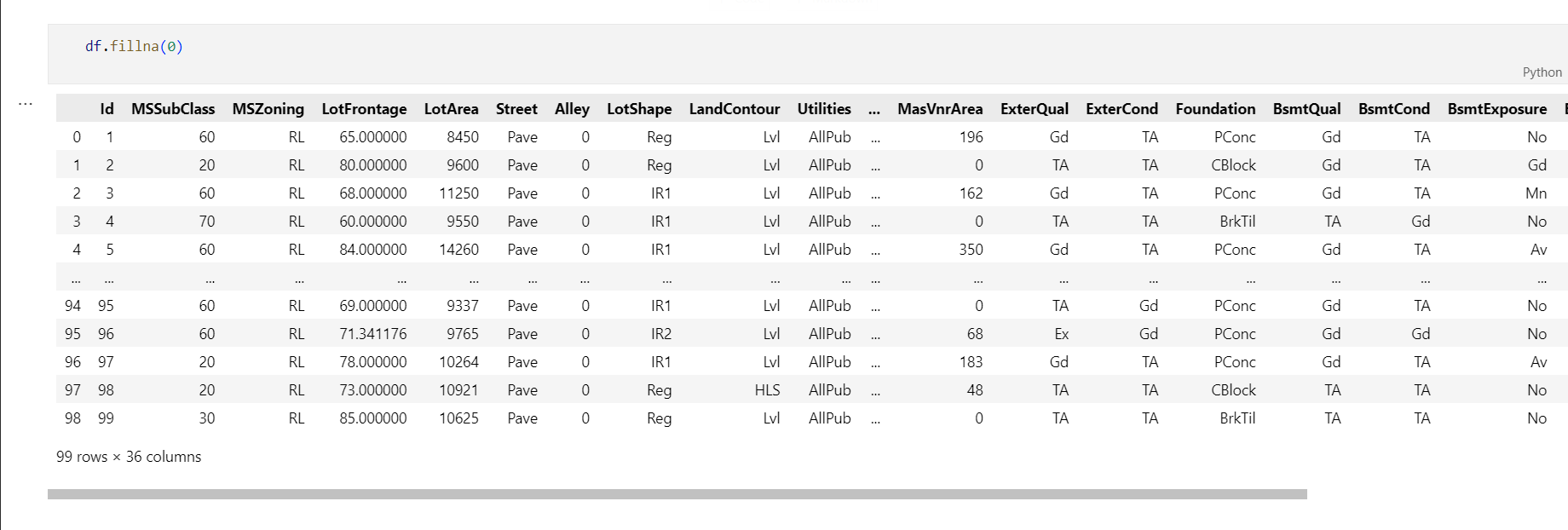


**2)Data formatting**

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.

**SOLN:**

****



#### 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':

**SOLN:**

# Create a list to store the 6 letter words

sixLetterWords= []

# Open the file

with open('about.txt') as fin:

    # Read each line

    for line in fin.readlines():

        # Split the line into words

        for word in line.split(" "):

            # Check each word's length

            if len(word) == 6:

                # Add the 6 letter word to the list

                sixLetterWords.append(word)

# Print out the result

print(sixLetterWords)

count = 0;

word = "";

maxCount = 0;

words = [];

#Opens a file in read mode

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

#Gets each line till end of file is reached

for line in file:

    #Splits each line into words

    string = line.lower().replace(',','').replace('.','').split(" ");

    #Adding all words generated in previous step into words

    for s in string:

        words.append(s);

#Determine the most repeated word in a file

for i in range(0, len(words)):

    count = 1;

    #Count each word in the file and store it in variable count

    for j in range(i+1, len(words)):

        if(words[i] == words[j]):

            count = count + 1;

    #If maxCount is less than count then store value of count in maxCount

    #and corresponding word to variable word

    if(count > maxCount):

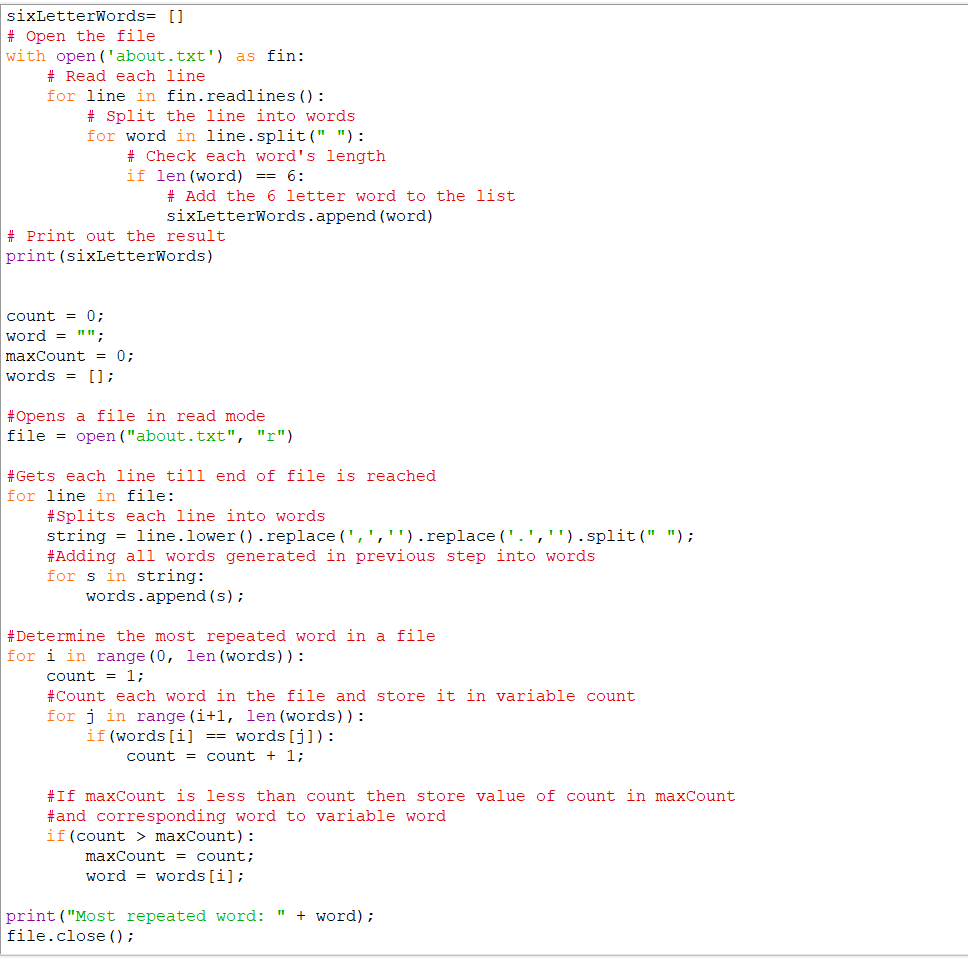
        maxCount = count;

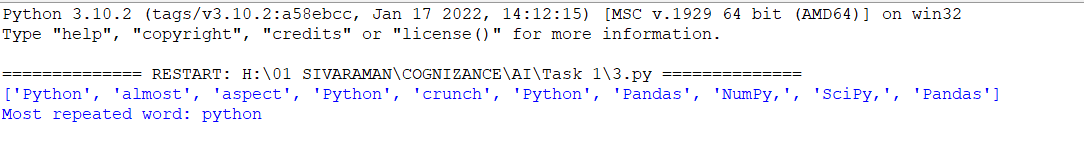
        word = words[i];

print("Most repeated word: " + word);

file.close();

**Screenshots:**

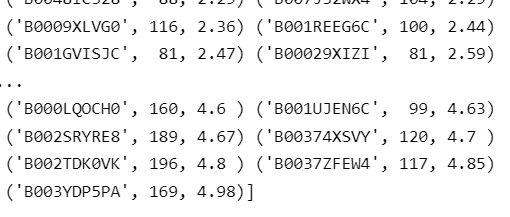
****

****

#### 4) Given food name, price, rating(out of 5) and their datatypes, create a structured array using NumPy and sort the array on rating

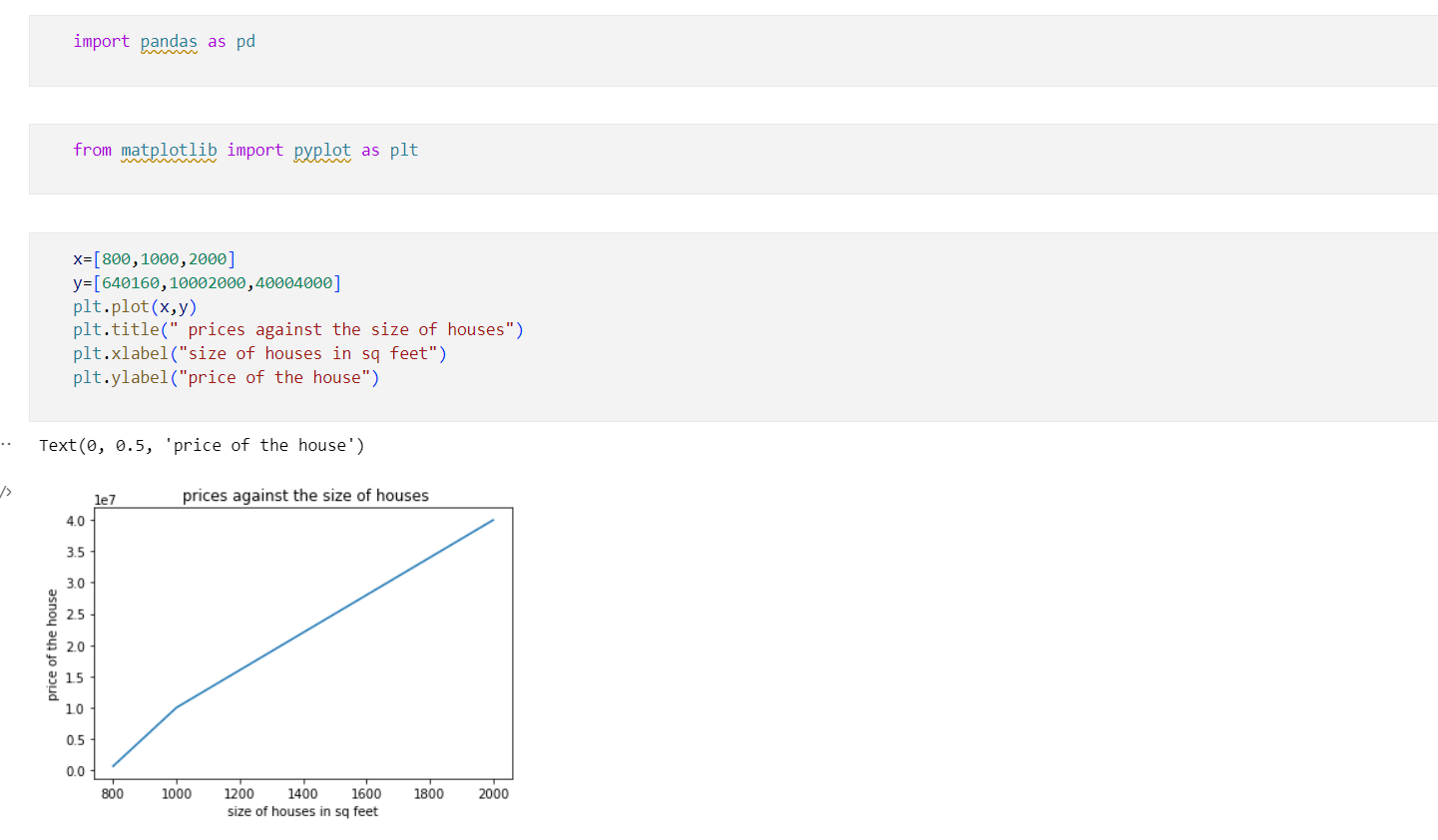
#### SOLN:

#### 



5) Let x be the size of houses in sq feet. x can take values between 700 to 2400 and is a multiple of 10. The price of the house (y) is found by the equation y = 10\*x^2 + 2\*x.

Pass the values into a .csv file and plot a graph of the prices against the size of houses. Provide appropriate labels for the axes.

****

