Python List Comprehensions and Regular Expressions (Pairs okay)

For this assignment, you will use Python to play detective. Write some list comprehensions and regular expressions to answer the questions below.

Part I: List Comprehensions

Begin with the following list definitions:

firstSentence = ['I', 'am', 'playing', 'xbox', 'and', 'trying', 'hard']

secondSentence = ['I', 'am', 'trying', 'hard', 'for', 'a', 'win']

Write the list comprehensions that compute the following lists. Here are the rules.

- Each question should be answered in one line of python, which is a list comprehension.

- Your code can only use numeric literals in the range -3 to 3.

- Each line can use no more than 2 string literals, totalling at most 3 letters long.

- If you are stuck, ask for a hint. The idea is to be able to write list comprehensions. Figuring out which one to write is just for fun.

1. ['playing', 'xbox', 'and', 'trying', 'hard', 'trying', 'hard', 'for', 'win'] (Has to do with word length)
2. ['am', 'and'] (has to do with first letters)
3. ['I', 'am', 'ng', 'ox', 'nd', 'ng', 'rd']
4. ['am', 'xbox', 'trying', 'hard', 'am', 'trying', 'hard'] (Has to do with word length)
5. [1,3,5,6] (Has to do with word length)
6. ['playing', 'trying', 'trying']
7. ['I', 'am'] (Has to do with comparing the lists)
8. ['I', 'am', 'trying', 'hard'] (Has to do with comparing the lists)
9. ['I', 'Am', 'Playing', 'Xbox', 'And', 'Trying', 'Hard']
10. ['I', 'eem', 'pleeying', 'xbox', 'eend', 'trying', 'heerd']
11. [('playing', 'trying'), ('xbox', 'hard'), ('and', 'for'), ('trying', 'a'), ('hard', 'win')] (Has to do with comparing the lists)
12. [('I', 'I'), ('am', 'am'), ('trying', 'trying'), ('hard', 'hard')] (Has to do with comparing the lists)
13. ['Iam', 'Itrying', 'Ihard', 'Ifor', 'Iwin', 'amtrying', 'amhard', 'amfor', 'amwin', 'xboxtrying', 'andtrying', 'andhard', 'hardtrying'] (Has to do with word length)
14. [0, 1, 5, 6] (Has to do with comparing the lists)
15. [('I', 'I'), ('I', 'a'), ('I', 't'), ('I', 'h'), ('I', 'f'), ('I', 'a'), ('I', 'w'), ('a', 'I'), ('a', 'a'), ('a', 't'), ('a', 'h'), ('a', 'f'), ('a', 'a'), ('a', 'w'), ('p', 'I'), ('p', 'a'), ('p', 't'), ('p', 'h'), ('p', 'f'), ('p', 'a'), ('p', 'w'), ('x', 'I'), ('x', 'a'), ('x', 't'), ('x', 'h'), ('x', 'f'), ('x', 'a'), ('x', 'w'), ('a', 'I'), ('a', 'a'), ('a', 't'), ('a', 'h'), ('a', 'f'), ('a', 'a'), ('a', 'w'), ('t', 'I'), ('t', 'a'), ('t', 't'), ('t', 'h'), ('t', 'f'), ('t', 'a'), ('t', 'w'), ('h', 'I'), ('h', 'a'), ('h', 't'), ('h', 'h'), ('h', 'f'), ('h', 'a'), ('h', 'w')]

For this part, print the lists to the screen along with the question number. Make sure the comments tell me exactly how to run this part.

Part II: Regular Expressions

For this part you will practice making regular expressions to find words that match patterns. The word list for you to use is here:

lowerwords.txt

This file contains more than 45,000 lower-case words, one word per line, in alphabetical order. Note that hyphens may be part of words. There are 12 questions about the words, and you will write a Python script that answers the questions.

The Questions:

1. How many words end with an 'a'?
2. How many five letter words are there that end with the letter 'd'?
3. How many words end with a vowel?
4. How many words start and end with a vowel?
5. How many words start and end with the same vowel, like "amoeba"?
6. The words "obsequious" and "pharmacopoeia" each contain 4 vowels in a row. How many words are there that contain four consecutive vowels?
7. The word "maintaining" contains the two-letter sequence "in" repeated three times. How many words are there that contain the sequence "in" three times (there may be other letters betweent the "in" repeats?
8. How many words contain some two-letter sequence repeated 3 times? For example, "contentment" and "maintaining" are such words because "contentment" has the sequence "nt" repeated three times and "maintaining" has the sequence "in" repeated three times.
9. The word "interpreter" contains the two-letter sequence "er" followed by its reverse "re" followed by the original "er". How many words are there with some two-letter sequence followed by the sequence reversed, followed by the original sequence again? (Letters between the repeats are ok.)
10. The word "bookkeeper" is famous for having three pairs of doubled letters in a row. If we allow letters to appear between the doubled letters, we get more matches, like "unsuccessfully" with its double-c, double-s, double-l. How many words have three pairs of doubled letters, not necessarily adjacent to one another?
11. Call a "triple" the pattern of one letter, followed by another, followed by the first, like "ama" or "ere". How many words have a triple that appears twice? For example, "amalgamate" has "ama" that appears twice, and "dereference" has "ere" that appears twice.How many words have three triples in a row? For example, "precipitateness" has "ipi" then "tat" then "ene" in a row.

Each question can be answered by filtering the list of words based on searching for a regular expression. You will need the following in your script:

A method called filterList that takes a regular expression and a list of strings as inputs and outputs a list that contains only words from the original list that match the regular expression. If you use a list comprehension, you can write this method in only one line.

The other tasks for the script can be done with whatever structure you like. In other words, you can make methods or not, with whatever arguments you want. Here is what needs to be done:

1. Read the words from the file into a list.
2. Call filterList once for each question.
3. Print the answers to the screen, along with their question number.

Note that the answer is always a number ("how many...") but you will probably want to print the actual words while testing to make sure you have the correct answer. Be sure to check not only that the words in the list fit the pattern, but also that you didn't leave any out.

Write comments and print statements so that it is clear how to run your script and what printed answer goes with which question.

Here is the rubric that will be used to grade you.