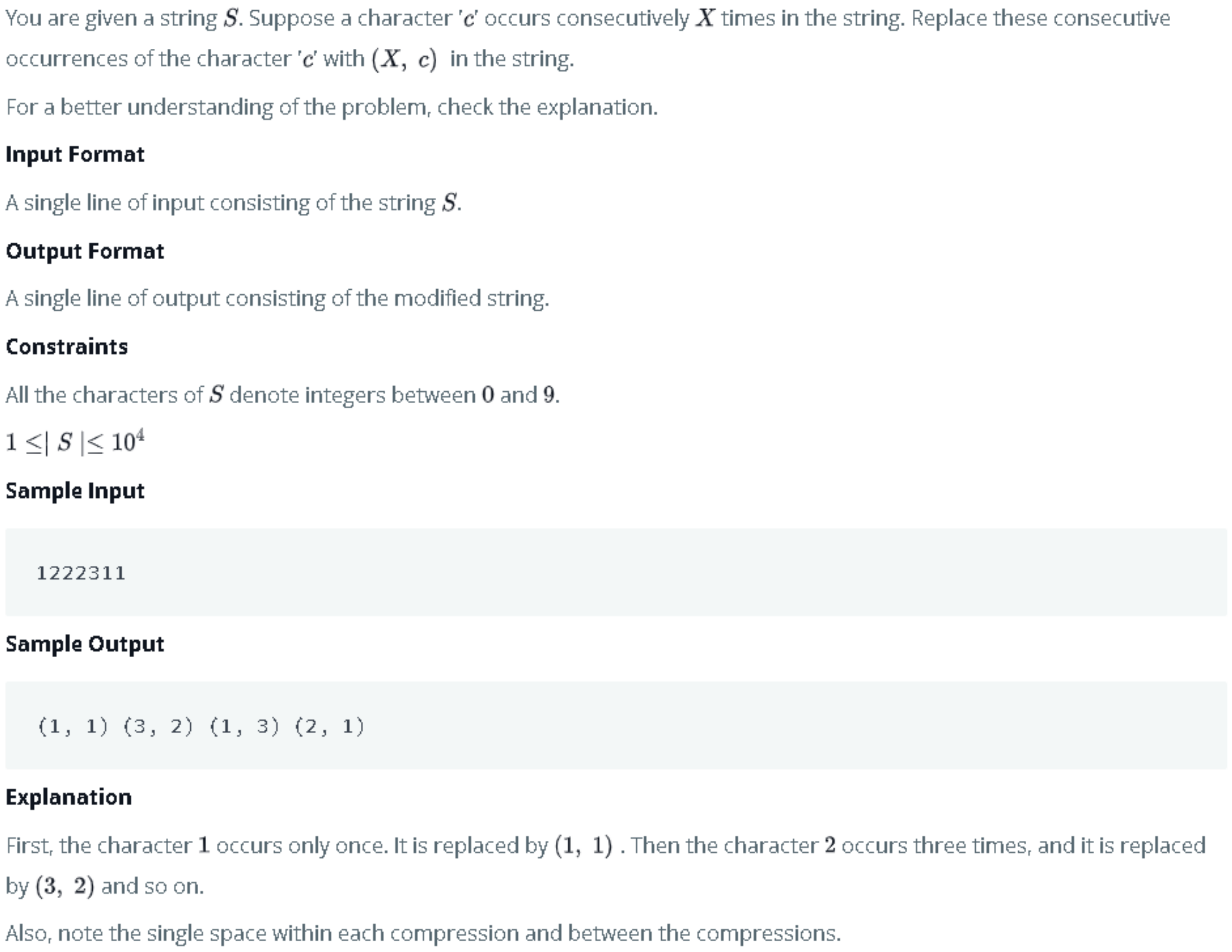
Problem 1: Simple Problem Solving



You are given a string S. Suppose a character 'c' occurs consecutively X times in the string. Replace these consecutive occurrences of the character 'c' with $(X,\ c)$ in the string.

For a better understanding of the problem, check the explanation.

Input Format

A single line of input consisting of the string S.

Output Format

A single line of output consisting of the modified string.

Constraints

All the characters of S denote integers between 0 and 9.

$1 \leq |\ S\ | \leq 10^4$

Sample Input

```
1222311
```

Sample Output

```
(1, 1) (3, 2) (1, 3) (2, 1)
```

Explanation

First, the character 1 occurs only once. It is replaced by $(1,\ 1)$. Then the character 2 occurs three times, and it is replaced by $(3,\ 2)$ and so on.

Also, note the single space within each compression and between the compressions.

Requirements:

- Solve in a **.py** File
- Comment your code
- Any code repetition will result in minus grades

Problem 2: Simple Bioinformatics Problem

1. Create a python script (.PY) called **fragment_lengths.py**

2. This script should calculate the **size of the two fragments** and **write them into a .txt file** that will be produced when the DNA sequence is digested with EcoRI

ACTGATCGATTACGTATAGTAGAATTCTATCATACATATATATCGATGCGTTCAT

Hint: The sequence contains a recognition site for the EcoRI restriction enzyme, which cuts at the motif G*AATTC (the position of the cut is indicated by an asterisk). Which means once the enzyme finds the sequence GAATTC, it will split the sequence after the **G** nucleotide base.

Requirements:

- Solve in a .py File or .ipynb
- Comment your code

Problem 3: Simple Image Processing Problem

You are required to get any **Colored** image and apply the following:

- Transform the image into **grey scale**
- Find the **inverse image** (Digital Negative) of the **grey scale** image
- Subplot the 3 images (colored, grey, inverse) in **1 Figure using matplotlib or seaborn.**

Requirements:

- Solve in a .py File or .ipynb
- Comment your code

Problem 4: Hard Bioinformatics/Data-science Problem

• In the materials folder, you'll find a text file called **data.csv**, containing some made-up data for a number of genes.

• Each line contains the following fields for a single gene in this order: species name, sequence, gene name, expression level.

Using **data.csv** to:

1. Print out the **gene names** for all genes belonging to Drosophila melanogaster or Drosophila simulans.

2. Print out the **gene names** for all genes between 90 and 110 bases long (sequence length).

3. Print out the **gene names** for all genes which has **AT content** (Search online what is AT content in Genes) is less than 0.5 and whose expression level is greater than 200.

4. Print out the gene names for all genes whose name begins with "k" or "h" **except** those belonging to Drosophila melanogaster.

Requirements:

- Solve in a (.ipynb) (Python Notebook)
- Each point must be done in a single cell
- You are **not allowed** to use anything except Pandas and Pandas conditions. (No if conditions and No loops)
- Comment your code
- Any code repetition will result in minus grades