

Data Science

1. **Google Colab** – This is a free service, based in *Jupyter Notebooks*, that supplies free GPU processing. It can be used, in a web browser, in the development of Python projects related to Computer Vision, Machine Learning and Data Science domains, using tools/libraries such as (Keras, PyTorch, TensorFlow,...).

Using one of the many resources (tutorials) available at the web (e.g., <https://towardsdatascience.com/getting-started-with-google-colab-f2fff97f594c>), the first task is to configure the Google Colab service and execute a simple “Hello World” notebook, using Python.

2. Create a function that translates emoticons to specific words. Given a sentence as a string, replace the emoticons ‘:D’, ‘:~’, ‘:(’ and ‘:P’ into the words ‘smile’, ‘grin’, ‘sad’ and ‘mad’.
3. Create a function that takes a string and returns a string in which each character is repeated once (example: "String" → "SSttrriinnngg")
4. Create a function that takes two numbers as arguments (num, length) and returns an array of multiples of num until the array length reaches length (example: (7, 5) → [7, 14, 21, 28, 35]).
5. In mathematics, a matrix (plural matrices) is a rectangular array or table of numbers, symbols, or expressions, arranged in rows and columns. In mathematics, a square matrix is a matrix with the same number of rows and columns. In linear algebra, the trace of a square matrix A is defined to be the sum of elements on the main diagonal (from the upper left to the lower right).

Example: A= [
 [0, 1, 0, 0],
 [1, 1, 1, 1],
 [0, 1, 0, 1],
 [0, 1, 1, 5]
]

$\text{tr}(A)=6$

6. In mathematics, primorial, denoted by “#”, is a function from natural numbers to natural numbers similar to the factorial function, but rather than successively

multiplying positive integers, the function only multiplies **prime numbers**. Create a function that takes an integer n and returns its **primorial**.

Examples: $\text{primordial}(2) = 6$ (the product of the first two prime numbers)

7. The number of rabbits banded at a series of sampling sites has been counted and entered into the following list. The first item in each sublist is an alphanumeric code for the site and the second value is the number of rabbits banded. Cut and paste the list into your assignment and then answer the following questions by printing them to the screen.

```
data = [['A1', 28], ['A2', 32], ['A3', 1], ['A4', 0],
        ['A5', 10], ['A6', 22], ['A7', 30], ['A8', 19],
        ['B1', 145], ['B2', 27], ['B3', 36], ['B4', 25],
        ['B5', 9], ['B6', 38], ['B7', 21], ['B8', 12],
        ['C1', 122], ['C2', 87], ['C3', 36], ['C4', 3],
        ['D1', 0], ['D2', 5], ['D3', 55], ['D4', 62],
        ['D5', 98], ['D6', 32]]
```

- a. How many sites are there?
 - b. How many rabbits were counted at the 7th site?
 - c. How many rabbits were counted at the last site?
 - d. What is the total number of rabbits counted across all sites?
 - e. What is the average number of rabbits seen on a site?
 - f. What is the total number of rabbits counted on sites with codes beginning with 'C'?
8. Suppose that you are interested in studying the relationship between the length of a specific species of animal ears and aspects of their DNA. A set of DNA samples is available, along with ear measurements from a small group of animals. The data are available at: <http://www.di.ubi.pt/~hugomcp/dataScience/dna.txt>

Write a Python script that:

- a. Imports the data into a data structure of your choice.
- b. Loops over the rows in the dataset.

- c. For each row in the dataset checks to see if the ear length is large (>10 cm) or small (≤ 10 cm) and determines the GC-content of the DNA sequence (i.e., the percentage of bases that are either G or C).
- d. Stores this information in a table where the first column has the ID for the individual, the second column contains the string 'large' or the string 'small' depending on the size of the individuals ears, and the third column contains the GC content of the DNA sequence.
- e. Prints the average GC-content for both large-eared animals and small-eared animals to the screen.