Data Science in Bioinformatics

Day 1

For all tasks please use text files for your answers and briefly comment your scripts within the code

1) **Basics**
   For refreshing your skills in the use of a terminal, create the following folder structure, without using any graphical UI. Which commands did you use?

```
| Master-Project-DS
      |- config
      |- data
            |- day1
      |- resources
      |- workflow
            |- envs
            |- rules
            |- scripts
```

2) **Concepts and Sequences**
   In bioinformatics it is all about biological sequences. Get an overview on the different types of concepts that are relevant for different biological processes and how they are connected:

   - DNA sequence
   - Protein sequence
   - Genotype
   - Protein biosynthesis

   DNA for example is a polymer composed of four different nucleobases (cytosine: C, guanine: G, adenine: A, thymine: T), encoding all genes that are necessary to produce different proteins. Please write a script that is capable of reading a random DNA sequence from the terminal command line:
   a. Read in from Terminal and output the sequence
   b. Transfer all the single letters in upper case
   c. Make sure the sequence only contains allowed letters (ACTG)
   d. Input a second DNA sequence and check if it's included in the first sequence

3) **File Handling**
   For analyzing large amounts of data, sequence data is stored in specific file formats. The most basic format is FASTA, a simple text file only containing plain text. Take a look at the provided file with an editor (VS Code, nano, vi) and name all the features you can spot.
   Write a script that is capable of reading the given file into a default python file structure. When running the script, it should provide a dialog presenting the available sequence names and outputting the sequence when inputting one of the given names.