

Scientific Data Engineer Challenge

Working as a Scientific Data Engineer at Exscientia will require that you are able to think on your feet and are able to both follow a design brief but also be able to propose innovative solutions of your own which the scientists might not have considered. This task is designed to give you the chance to prove your skills as an innovator and you are free to build whatever application you would like, the only rule is that your application must be written in Python.

As a rough guide we would suggest you spend between 2 - 4 hours on the challenge.

Challenge:

Write a program that takes an SD file and returns properties of the molecule(s) such as:

- MW
- SMILES
- IUPAC name
- Image
- Any examples of your own

A component of the program should be able to understand and compare properties of a set of molecules in the file, for example:

- sort compounds by launched date
- sort compounds by therapeutic class, and within each therapeutic class by molecular weight
- something you think may be useful and interesting!

Please include unit tests - the attached example SDF may be useful to use. Don't be afraid to over engineer your tests or have duplicates to showcase your skills and different ways to approach the problem.