Automation with Python

Y ou may have heard that scripting languages, such as Python, Perl, BASH and R, are well-suited for 'automating the boring tasks'. Hopefully you will see this is true here!

In this section, we will briefly examine how to automate a repetitive work-flow using Python. We will work with healthcare data. The Human Resources (HR) department would like an Excel spread-sheet of whom worked on which patient visit, so they can send them cookies and/or flowers as a bonus.

They would like to do this every month, so you don't want to have to do this 'manually' every time, do you?

Data

The catch here is that the visit and patient data is stored on the hospital's SQL database, but the detailed staff information is stored by HR as an Excel spreadsheet. Respectively, these are:

`data/healthcare/Staff.xlsx`

`data/healthcare/healthcare.sql`

We'll need to combine these data.

Unlike other concepts we've covered so far, the resulting product of this exercise will be a stand-alone script -- a file containing the code -- that can be run automatically each time.

---

1. Access the 'visit' data from the hospital's SQL database. For simplicity, this is a SQLite database (the most widely used db in the world!)
2. Open a database connection and obtain the data from one of the tables within (check the table names)
3. Read in the contents of HR's `Staff.xlsx`
4. HR wanted to know the staff working on each visit. We'll need to join the staff information to the visit data. What kind of merge/join operation would this be?
5. Output the result as an Excel spreadsheet
6. Package the code in a format that can be run externally
7. Apply some kind of scheduler to run this script each month. For example, on a Windows desktop, this could be `Task Scheduler`, or `cron` for Linux/MacOS