

Overview:

We would like you to complete a task, demonstrating your ability to manipulate, cleanse and analyse data using Python, SQL and data visualisation tools. Below, we have provided you with an explanation of the data source we would like you to use, some instructions on the tasks we want you to perform, and some additional considerations.

Explanation of Data:

- The following site contains eight CSV files, updated weekly, by each local UK Traffic Commissioner authority... <https://www.data.gov.uk/dataset/2a67d1ee-8f1b-43a3-8bc6-e8772d162a3c/traffic-commissioners-goods-and-public-service-vehicle-operator-licencerecords>
- *Note - please ignore the last updated date on this page, as it is incorrect.*
- The data contains licencing information for Goods Vehicles and Public Service Vehicles.
- The licence number shows what type of licence each one is:
- Licence number beginning with 'O' - Goods Vehicle Licence
- Licence number beginning with 'P' - Public Service Vehicle Licence
- More information on vehicle licencing can be found here:
- HGV - <https://www.gov.uk/being-a-goods-vehicle-operator>
- PSV - <https://www.gov.uk/psv-operator-licences>
- Each of the eight local traffic authorities can issue licences, the licence number can be found in column B in each of the spreadsheets.
- Licence data has been duplicated, as the data contains an individual row for each Transport Manager (column H) and/or Company Director (column M).
- Some of the data also contains the company registration number - used to identify a limited company with companies house:
- Link to companies house - <https://www.gov.uk/get-information-about-a-company>.
- Further data can be pulled using the company registration number via the companies house API - <https://developer.company-information.service.gov.uk/overview>
- Companies may hold one or more licences in each traffic authority.

Instructions:

- Use Python to create an automated process to download, extract and manipulate the CSV files.
- Combine into a single dataset, and store in a structured database.
- Cleanse data to remove any duplicates, providing a dataset showing each licence, and a dataset showing each registered company.
- Produce analysis using a tool of your choice.
- Provide some high-level analysis on the licence data, and company data. Focusing on the number of Vehicles Specified (column K).