# Extract, Transform, Load

# Book Store

In this project I aim to create a database of books purchased from an online retailer using 3 data sets ‘Books’, ‘Ratings’ and ‘Users’.

The goal is to end up with a complete record for each instance in the final table; complete with information about the buyer and the product by combining the three datasets.

I will use the ETL process to create my database by first sourcing the datasets from an open-source website such as Kaggle, cleaning and preparing the data using Excel and Python, and finally using relational database tools to create the final database.

# Tools and Methods

## Excel

Use Excel to interrogate the datasets and carry out any necessary data preparation steps

## Python and the Pandas library

Use Python tools within Jupyter Notebook to carry out data cleaning such as renaming and re-arranging of columns, removing null values etc

## Pg Admin

Will be used to create the SQL database and load the final data frame ready for querying

## SQLalchemy

Will be used to carry out the necessary Object Relational Mapping and allow for creating the relational database alongside Python.

# Data sources

The data source contains 3 data sets from an online retailer and contains books along with user ratings and some information about the customers who bought them.

<https://www.kaggle.com/datasets/saurabhbagchi/books-dataset>