# Module 2

# Online retail

This assignment will address the main skills for cleaning and tidying your datasets like

## Looking at and cleaning datasets

## Selecting and filtering datasets

## Arranging datasets

## Adding new variables or columns

## Creating useful summaries

This is a transnational data set which contains all the transactions for a UK-based and registered non-store online retail. The company mainly sells unique all-occasion gifts. Many customers of the company are wholesalers. The data set can be downloaded from:

<https://archive.ics.uci.edu/ml/machine-learning-databases/00352/Online%20Retail.xlsx>

See also

Daqing Chen, Sai Liang Sain, and Kun Guo, Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining, Journal of Database Marketing and Customer Strategy Management, Vol. 19, No. 3, pp. 197â€“208, 2012 (Published online before print: 27 August 2012. doi: 10.1057/dbm.2012.17).

# Description of the dataset

A data frame of 541909 rows (transactions) on 8 columns (variables):

1. InvoiceNo: Invoice number, a unique value for each transaction. Codes start with `c` indicate cancellation.
2. StockCode: Product code, unique for each product
3. Description: Product Name
4. Quantity: of each product per transaction
5. InvoiceDate: date when each transaction was generated
6. UnitPrice: product price per unit in sterling
7. CustomerID: customer number, unique for each customer
8. Country: country name where each customer resides

# Questions

1. Create a subset of this dataframe that contains all the observations but with no missing values. How many rows does this subset contain?
2. Create another subset of the original dataframe that contains unit prices greater than 1000 and ‘United Kingdom’ and ‘Australia’ countries. How many rows does this subset contain?
3. What is the earliest and latest dates of this dataframe?
4. Add another column, called ‘total’= Quantity \* UnitPrice. Which product name has the highest mean total? What is its code?
5. Which weekday has the highest mean total? Which country has the highest maximum total?