Group coursework 4

- Please submit your coursework on Moodle **by Midday on 16/04**.

- Please upload your reports to Questions 1 and 2 in **one pdf file** with no more than 15 pages long including plots, of which no more than 6 pages should be text.

- Please also upload one R script for Question 1 and one R script for Question 2.

- This coursework will be assessed based on the quality of your report, e.g. presentation, quality of plots, details about the experiment settings, motivations to use the methods, analysis of the results, conclusions etc.

- Make sure that you have included **sufficient** comments in the codes to make them **readable** by other people. There should be **no error messages** shown when I run your R scripts. You can assume that I have installed all required packages.

Question 1 [**20 marks**]

Download the `Online_Retail.xlsx` data from Moodle. This is a transnational dataset which contains all the transactions occurring between 01/12/2010 and 09/12/2011 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.

Explore the dataset and report any interesting findings.

Question 2 [**20 marks**]

Download the `US-Stock.cvs` data from Moodle. This dataset collects financial indicators for the stocks of the US stock market in 2018. The last column of the dataset represent the class of each stock: if the percent price variation of the stock for the year is positive, then `class=1`; otherwise, `class=0`.

Explore the dataset, compare different classification methods and report your findings.