

## **COMP 4112 Introduction to Data Science**

### **Assignment 3, Unsupervised learning**

**5 marks, 5 percent total**

**Submission:**

**Submit a single .py file named “<username>\_A3.py” on myCourselink.**

**Grading:**

**1/5 – Does not function**

**2/5 – Partially functioning**

**3/5 – Basic features, no evidence of experimentation**

**4 and 5 / 5 – Working and evidence of some experimentation**

In this assignment you will cluster a total of 2 datasets we have looked at in the course. You can select the datasets you want to cluster. Additionally, try and report some information regarding the clustering. There are no specific requirements for this assignment other than clustering of 2 chosen datasets; the amount of detail you want to report about the clustering is up to you, but ideally it should show some evidence of effort and experimentation.

**For this assignment (Python only):**

**Read in the CSV datasets. You can do this how you like; Python lists are totally acceptable but you might have to convert to other formats for scikit-learn sometimes. If you want, you could use a pandas dataframe or a numpy array.**

**Other Hints:**

**Code examples provided can be modified to assist with completing this assignment.**