

**Department of Computing,
The University of the West Indies**

COMP6115 – Knowledge Discovery and Data Analytics I

Assignment 2 – Unsupervised Learning [20 marks]

Due April 30, 2022

Task 01 - Clustering [10 marks]

Using the `SalesData` which represents regional purchases of different consumer products over a few years, complete tasks 1-3.

1. Clean and prepare the data appropriately and conduct the clustering analysis.
 - a. Apply binning and normalization as necessary to improve results.
 - b. Construct new field which has the number of days between order Date and Ship Date.
 - c. Provide short in-code comments to explain the reason and what choices are made.
2. Generate 4, 6, and 8 cluster results from the dataset in R, using both `hierarchical clustering` and `k-means` (6 separate models).
3. Review and visualize the clusters and determine what is observed in each result (write short notes about your observations).

Task 02 – Association Rule Mining [10 marks]

You have been provided data on an International online retail outlet for several countries in the file `<online retail.csv>`. In class we had extracted association rules for the country “France” from this data set for a given Minimum Support and Confidence thresholds. In this assignment for the country that is allocated to you in the table below, identify association rules that can be used by the country manager to understand the buying patterns of their customers. You will have to identify the settings for Minimum Support and Confidence and present the extracted association rules. Comment your code and provide visualizations for the manager to explore different rules and their associated support, confidence and lift values.

Student Name	Country	Number of Records
HOPE, Bertland	Netherlands	2371
HAMILTON, Sanjay	Germany	9495
NISBETT, Cheryl	Switzerland	2002
LAWRENCE, Ackeem	Spain	2533
SENIOR, Ongel	Portugal	1519
DUCKIE, Owen	EIRE (Ireland)	8196
HARRIS, Donielle	Norway	1086
PLUMMER, Navion	Belgium	2069
BARNES, Rushane	Australia	1259
GRAHAM, Toni- Ann	Netherlands	2371
GREEN, Adrian	Germany	9495

MILLS, Ottor	Switzerland	2002
KEITH, Devon	Spain	2533
CRUICKSHANK, Jamelia	Portugal	1519
PALMER, Aldon	EIRE (Ireland)	8196
THOMPSON, Bradley	Norway	1086

Submission:

Submit a single file with your R code and notes for both tasks (you may use ***rmarkdown*** to submit a pdf showing code, visualizations, and notes).