

CONFIDENTIAL



UNIVERSITI TUN HUSSEIN ONN MALAYSIA

**FINAL EXAMINATION
SEMESTER I
SESSION 2019/2020**

COURSE NAME : ADVANCED DATA MINING
COURSE CODE : BWB 43403
PROGRAMME CODE : BWQ
EXAMINATION DATE : DECEMBER 2019 / JANUARY 2020
DURATION : 3 HOURS
INSTRUCTION : ANSWER ALL QUESTIONS

THIS QUESTION PAPER CONSISTS OF TEN (10) PAGES

TERBUKA

CONFIDENTIAL

- Q1** (a) Dataset can be identified in three types which are structure, un-structure and semi-structure. Elaborate in the details **TWO (2)** advantages and **TWO (2)** disadvantages between these types of data. (8 marks)
- (b) Based on your answers in **Q1(a)**, provide **TWO (2)** relevant examples or usage of these types of data in e-Businesses. (6 marks)
- (c) Differentiate between nominal, ordinal, interval and ratio attributes. Give **ONE (1)** example for each attribute. (8 marks)
- Q2** (a) Suppose that the data for analysis includes the attribute age. The age values for the data tuples are presented in **Table Q2(b)**. Answer the following questions.

Table Q2(b) : Age

13	21
15	22
23	30
16	25
19	25
23	28
20	35

- (i) Transform the data using min-max normalization. (8 marks)
- (ii) Transform the data by using Z-score normalization. (9 marks)
- (iii) Comment on which method you would prefer to use and provide the reasons. (4 marks)
- (b) Analytics spectrum can be classified into four categories. Based on this information, answers the following questions.
- (i) List all **FOUR (4)** categories. (4 marks)
- (ii) Explain briefly for each category. (4 marks)
- (iii) Suggest the suitable technique to be used in each category. (4 marks)

TERBUKA

- Q3** (a) Discuss the use of *Support* and *Confidence* in association rules. (4 marks)
- (b) **Figure Q3(b)** shows the transactions from the market basket analysis in Supermarket Value. Give **THREE (3)** interpretations from this transactions. (3 marks)
- (c) From **Figure Q3(c)**, summarizes **SIX (6)** of the market basket analysis. (6 marks)
- Q4** (a) K-means is one of the several data mining techniques. Discuss **TWO (2)** issues or limitations of this technique. (4 marks)
- (b) **Figure Q4(b)** shows the output of k-means clustering. List all the cities with their groups. (6 marks)
- (c) Based on **Figure Q4(c)**, if the height of the *cutree* is four, calculate the number of clusters can be formed. Justify your answers and give the total number of each of the cluster. (4 marks)
- (d) **Figure Q4(d)** shows the output from **three** clustering methods to find the best number of clusters. Summarizes **TWO (2)** similarities and **TWO (2)** dissimilarities between these plots. (8 marks)
- (e) Give **THREE (3)** summarizations from **Figure Q4(e)**. (6 marks)
- (f) From your answers in **Q4(b)-Q4(e)**, conclude **TWO (2)** similarities and **TWO (2)** dissimilarities from k-means and hierarchical clustering approaches. (4 marks)

– END OF QUESTIONS –

TERBUKA

FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2019/2020
COURSE NAME : TECHNIQUES OF DATA MINING

PROGRAMME CODE : BWQ
COURSE CODE : BWB 44603

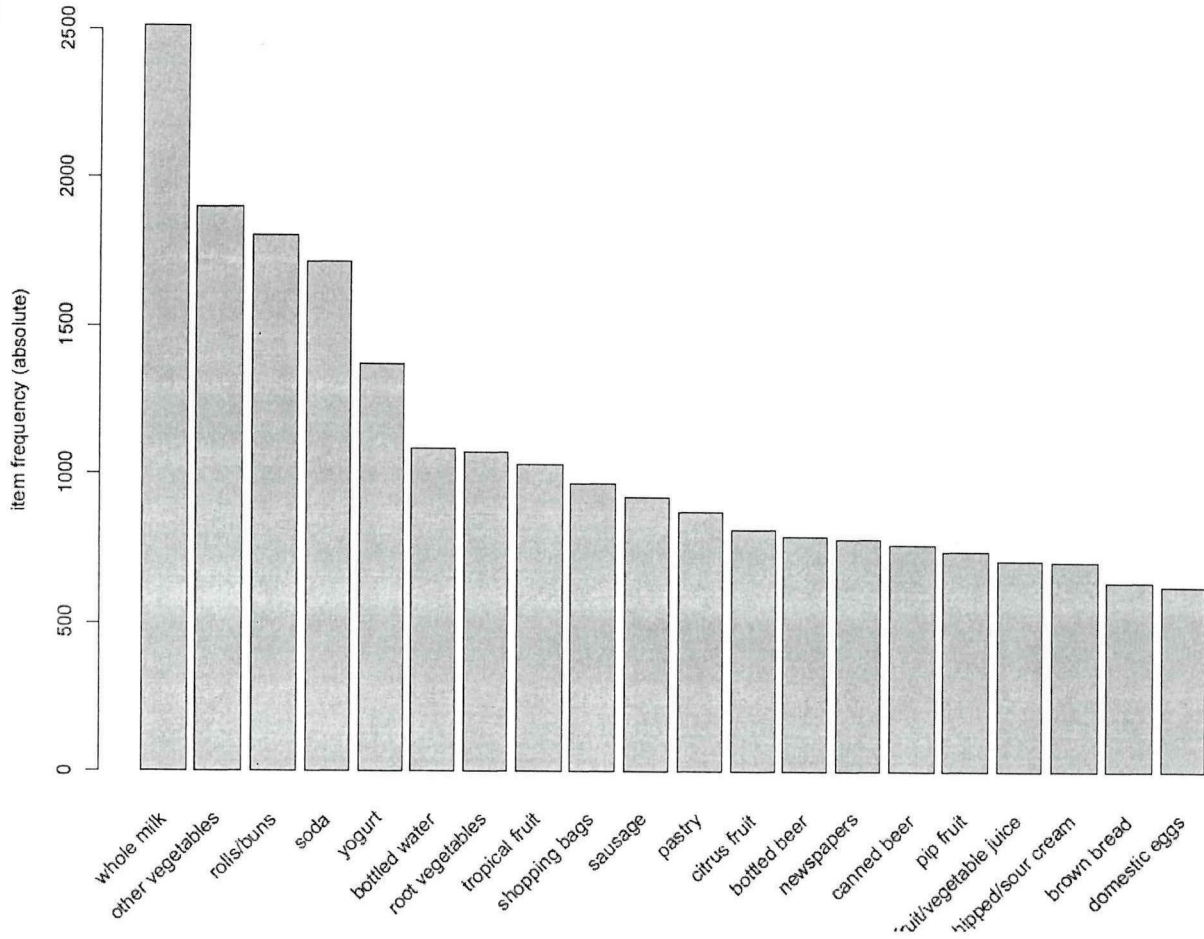


Figure Q3(b)

TERBUKA

FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2019/2020
COURSE NAME : TECHNIQUES OF DATA MINING

PROGRAMME CODE : BWQ
COURSE CODE : BWB 44603

```
> inspect(rules[1:5])
  lhs                rhs                support confidence lift count
[1] {rice,           => {whole milk}  0.0012           1 3.9 12
    sugar}
[2] {canned fish,   => {whole milk}  0.0011           1 3.9 11
    hygiene articles}
[3] {root vegetables, butter,
    rice}           => {whole milk}  0.0010           1 3.9 10
[4] {root vegetables, whipped/sour cream,
    flour}          => {whole milk}  0.0017           1 3.9 17
[5] {butter,        => {whole milk}  0.0010           1 3.9 10
    soft cheese,
    domestic eggs}
```

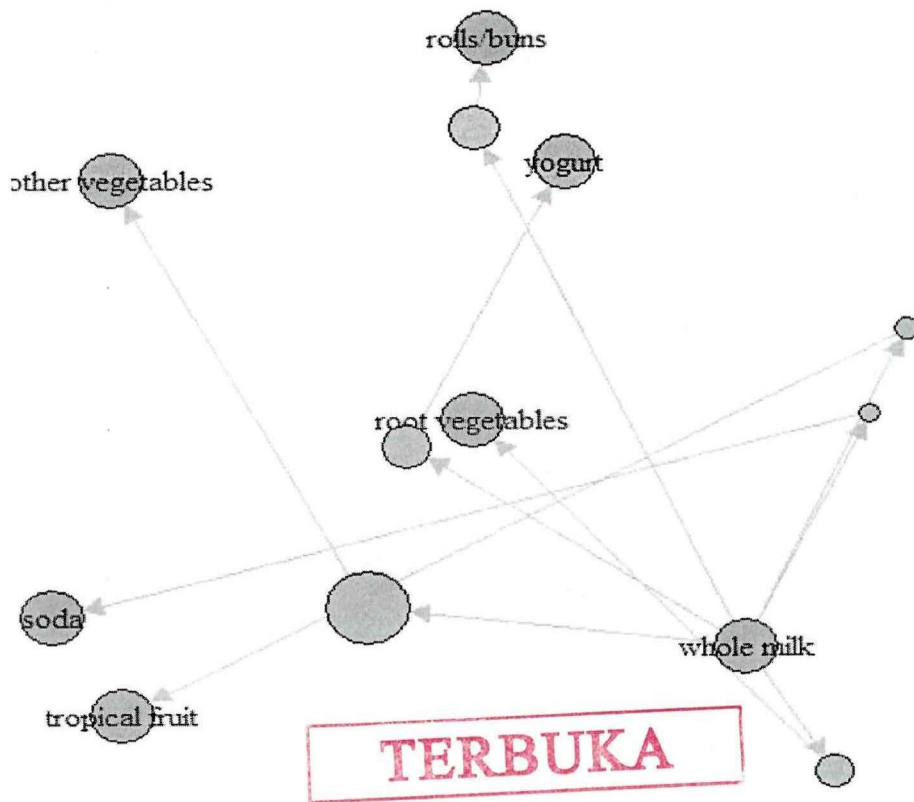


Figure Q3(c)

FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2019/2020
 COURSE NAME : ADVANCED DATA MINING

PROGRAMME CODE : BWQ
 COURSE CODE : BWB 43403

Clustering vector:

Alabama	Alaska	Arizona	Arkansas	California
4	3	3	4	3
Colorado	Connecticut	Delaware	Florida	Georgia
3	2	2	3	4
Hawaii	Idaho	Illinois	Indiana	Iowa
2	1	3	2	1
Kansas	Kentucky	Louisiana	Maine	Maryland
2	1	4	1	3
Massachusetts	Michigan	Minnesota	Mississippi	Missouri
2	3	1	4	3
Montana	Nebraska	Nevada	New Hampshire	New Jersey
1	1	3	1	2
New Mexico	New York	North Carolina	North Dakota	Ohio
3	3	4	1	2
Oklahoma	Oregon	Pennsylvania	Rhode Island	South Carolina
2	2	2	2	4
South Dakota	Tennessee	Texas	Utah	Vermont
1	4	3	2	1
Virginia	Washington	West Virginia	Wisconsin	Wyoming
2	2	1	1	2

within cluster sum of squares by cluster:

[1] 12.0 16.2 19.9 8.3
 (between_ss / total_ss = 71.2 %)

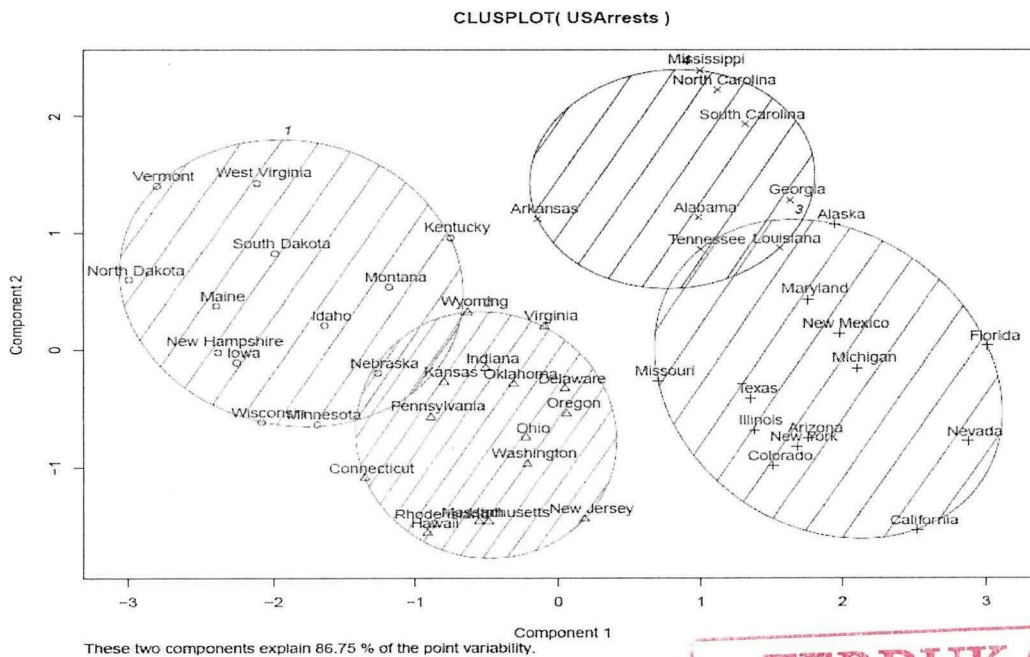
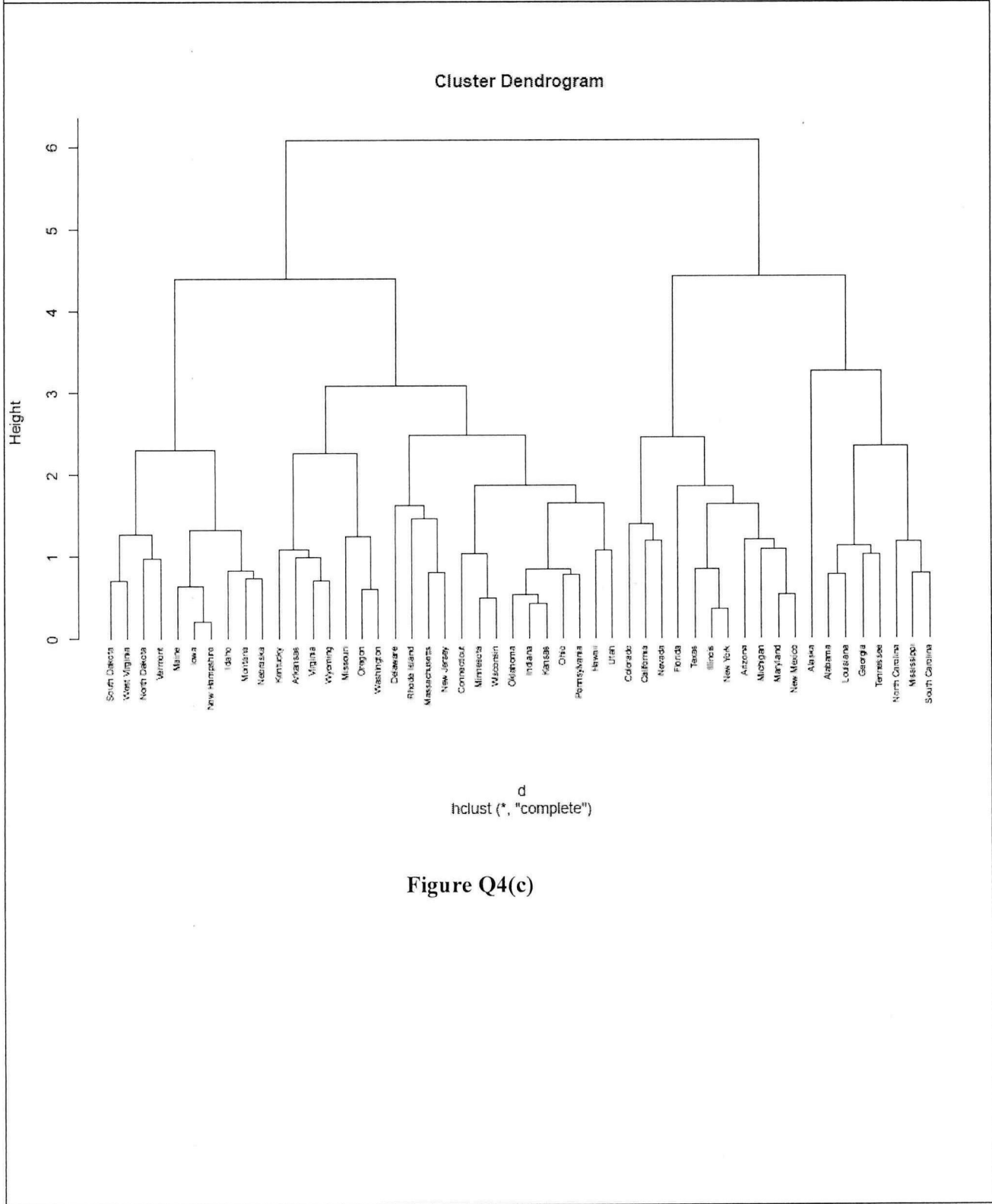


Figure Q4(b)

FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2019/2020
COURSE NAME : ADVANCED DATA MINING

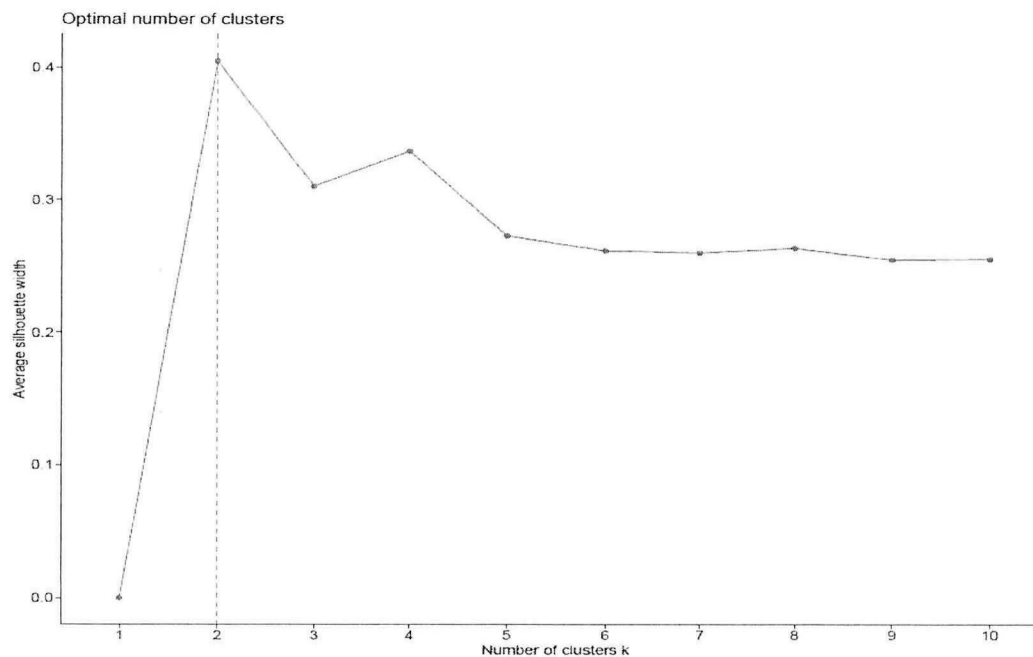
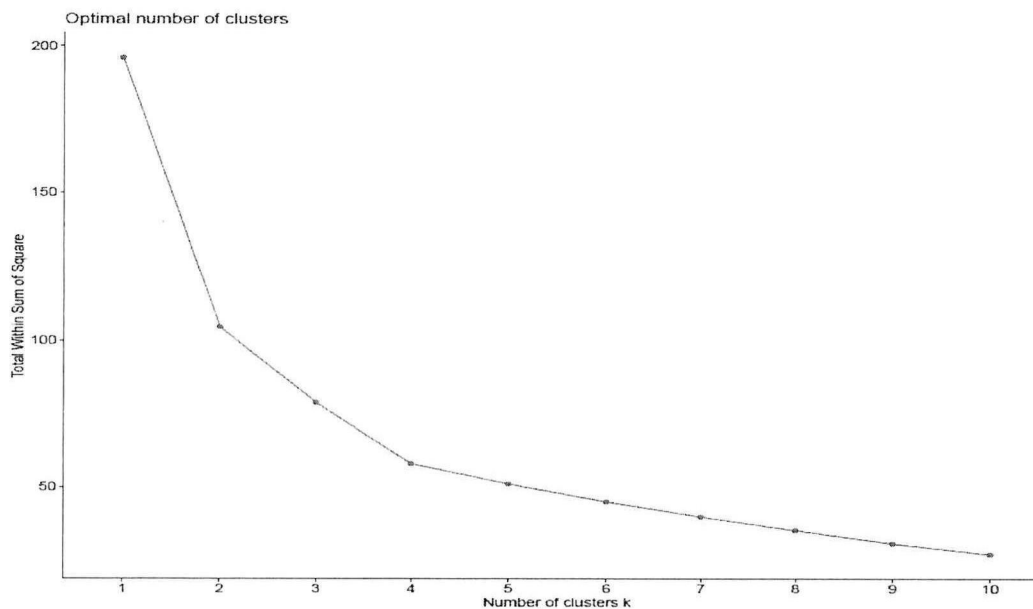
PROGRAMME CODE : BWQ
COURSE CODE : BWB 43403



FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2019/2020
COURSE NAME : ADVANCED DATA MINING

PROGRAMME CODE : BWQ
COURSE CODE : BWB 43403



TERBUKA

FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2019/2020
COURSE NAME : ADVANCED DATA MINING

PROGRAMME CODE : BWQ
COURSE CODE : BWB 43403

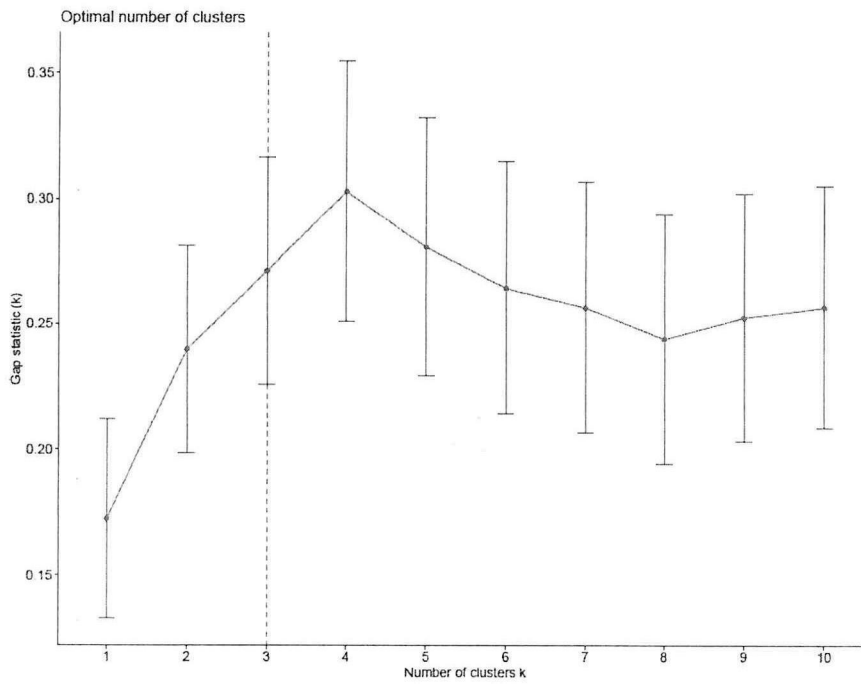


Figure Q4(d)

TERBUKA

FINAL EXAMINATION

SEMESTER / SESSION : SEM I / 2019/2020
COURSE NAME : ADVANCED DATA MINING

PROGRAMME CODE : BWQ
COURSE CODE : BWB 43403

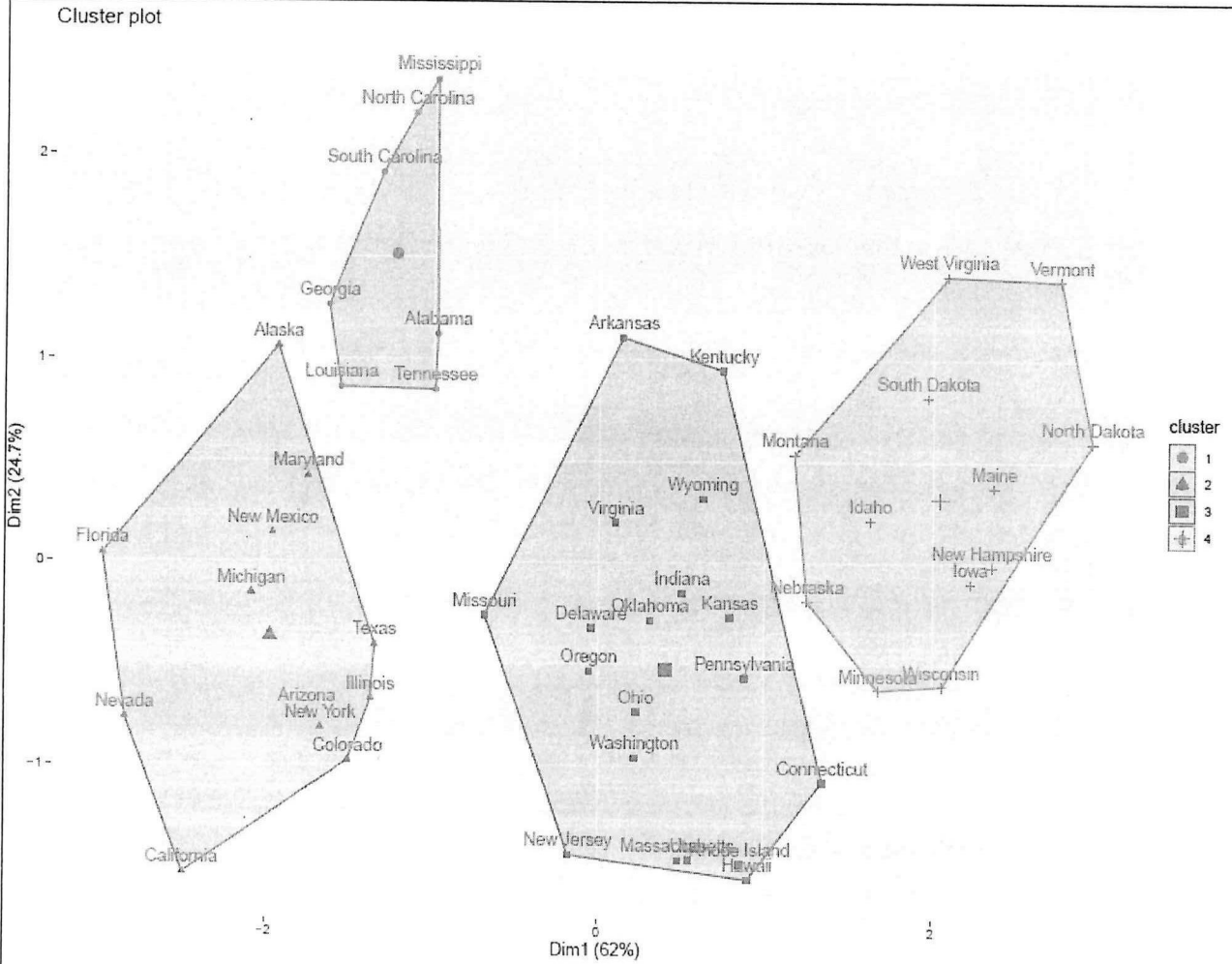


Figure Q4(e)

TERBUKA