C11BD Big Data Analytics: Individual Coursework 2

You are working for a grocery company, FreshCo, that wishes to segment their customers in order to target their marketing more effectively. You have been given a sample of customer data.

Perform an exploratory data analysis on the data and then use k-means to produce a cluster analysis. Prepare a report documenting your methodology, results and interpretations. The report should include the following:

- Description of the data preparation (including your R code in appendix) with an exploratory data analysis noting any data quality issues and how it is managed
- Description of how the clustering is performed (including your R code in appendix)
- Results of the clustering performed visualised
- Discussion on the findings, considering the usefulness and limitations to the company

Dataset

Explore the dataset containing information about a customer subset of FreshCo. The dataset includes 2093 records/instances (each representing a customer) with seven features describing the customer. The dataset, *dataset_FreshCo.csv*, is provided accompanied by a text file, *dataset_FreshCo_CodeBook.txt*, that provides a description of each feature.

Instructions

- Complete the assignment individually.
- Your final report should be a maximum of 3,000 words. It should include an executive summary, an introduction, body, conclusion, and, if necessary, a reference list. Note that the executive summary , reference list and R code does not form part of the word count.
- Your R code should be distinguished in the text using `Courier New` font and provided as an appendix to the report – be sure to make use of comments.
- You need to include your **certificate for completing the LinkedIn Learning course** on *Learning R* in an appendix to the report.
- The assignment is due in Week 10 on **16ᵗʰ March before 12pm (GMT) / 4pm (Dubai).**

The assignment will account for 60% of your final course mark.

The report will be marked according to the following criteria:

- Approach and methodology
- Analysis of data and results
- Interpretation of findings
- Appropriate professional writing and presentation style
- Certificate of LinkedIn Learning course

Supporting Resources

Refer to the prescribed textbooks, lecture slides, tutorials and LinkedIn Learning course. You are further encouraged to search through other resources as necessary.