Chow Hound: Food for Thought

Case questions require the following auxiliary data files, available on Brightspace.

- *DSO528_clustering_case.csv*

Case Questions:

1. Create a new variable Time for the membership length in years for all customers.

2. Using all variables (including Time) EXCEPT **JoinDate**, form K-means clusters for clusters between 3 and 10. Based on Silhouette analysis, how many clusters are suggested to form? Report the Silhouette score. Justify your response quantitatively.

3. Next, choose 6 clusters and reproduce the best clustering found in the previous part. Briefly describe the common characteristics for each segment of customers. You may give each segment a shorthand name to help with the illustration.
   Hint: Visuals such as coordinate plots and biplot may help explain.

4. Based on your analysis, do you agree with Harry's claim that most of Chow Hounds customers are college students? Justify your response.

5. Based on your analysis in the previous part, which clusters do you think are most significant/valuable to help the business grow? If you think that this depends on the circumstance, please give examples and address your assumptions.

6. Next, choose 7 clusters and reproduce the best clustering found in the part (2). What is the main difference between the 6 clusters formed in the previous question? From the management standpoint, which solution would you propose? Briefly explain.

7. We included "Spending" in the previous analysis, but "Spending" is the sum of the cuisine amounts across different variables. Some may argue that this results in "double counting" the effect of spending. Refit the K-means model using the same variables but exclude "Spending." Compare and contrast the resulting segmentation with the one from Question 2, both qualitatively and quantitatively. Are there any significant differences? If so, which model would you recommend?
   Note: For comparison purposes, set *n=6* and *n_init=100* to find the best clustering without "Spending."

8. Joan is proposing moving from their current pricing (a flat percentage of 10% on all transactions) to a flat charge of $1.50 per order. How much money do you expect to gain/lose in each of the clusters identified in part 2 under the new pricing? Which pricing scheme will you suggest?
   (Hint: use Excel pivot table.)