**BAN 5743: Individual Exercise 1 (10 Points)**

**Problem Introduction and Data Description**

Elements of the data dictionary as shown below describe survey data on people’s attitudes and opinions, and questions concerning oral hygiene. A large dental non-profit organization is wanting to do a nationwide campaign to improve dental outcomes. In order to have an effective campaign, the marketing department is asking for your help to analyze and cluster the survey data.

Software: SAS Enterprise Miner

Dataset: conspref.sas7bdat

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Description | Dictionary | Type |
| Response\_ID | Response ID |  | ID |
| Gender | Gender | Male, Female | Descriptor |
| Birth\_Year | Birth Year |  | Descriptor |
| Single\_Status | Single Status | 1 = Single, 2=Not Single | Descriptor |
| School\_Age\_Children | School Age Children | 1 = Yes, 2 = No | Descriptor |
| Age\_in\_Years | Age in Years | Continuous | Descriptor |
| YRS\_Current\_Employer | Years at Current Employer | Continuous | Descriptor |
| YRS\_Current\_Position | Years in Current Position | Continuous | Descriptor |
| Salary | Salary | Continuous | Descriptor |
| Job\_Satisfaction | Job Satisfaction | 1 = Not at all satisfied,  2 = Somewhat satisfied,  3 = Extremely satisfied | Descriptor |
| career | I am working on my career | 1 = Agree, 2 = Disagree | Descriptor |
| world | I want to see the world | 1 = Agree, 2 = Disagree | Descriptor |
| homeimprov | My home needs some major improvements | 1 = Agree, 2 = Disagree | Descriptor |
| outsidework | I have vast interests outside of work | 1 = Agree, 2 = Disagree | Descriptor |
| debtundercontrol | I want to get my debt under control | 1 = Agree, 2 = Disagree | Descriptor |
| largefamily | I come from a large family | 1 = Agree, 2 = Disagree | Descriptor |
| Brush | Brush | 1 = Almost Every Day,  2 = Every Day,  3 = Twice a day,  4 = More than twice a day | Base |
| Floss | Floss | 1 = Rarely or Never floss,  2 = Almost as often as I brush,  3 = Every time I brush my teeth,  4 = More often than I brush | Base |
| B\_wakingup | Brush After Waking Up | 1 = Yes, 2 = No | Base |
| B\_aftermeal | Brush After Meal | 1 = Yes, 2 = No | Base |
| B\_beforesleep | Brush Before Sleep | 1 = Yes, 2 = No | Base |
| B\_anothertime | Brush Another Time | 1 = Yes, 2 = No | Base |
| B\_other | Brush Other | 1 = Yes, 2 = No | Base |
| F\_wakingup | Floss After Waking Up | 1 = Yes, 2 = No | Base |
| F\_aftermeal | Floss After Meal | 1 = Yes, 2 = No | Base |
| F\_beforesleep | Floss Before Sleep | 1 = Yes, 2 = No | Base |
| F\_anothertime | Floss Another Time | 1 = Yes, 2 = No | Base |
| F\_other | Floss Other | 1 = Yes, 2 = No | Base |
| Freq\_teethcleaning | Frequency of Teeth Cleaning | Continuous | Base |
| num\_fillings | Number of Fillings | Continuous | Base |
| TP\_Cost | Toothpaste Cost | Continuous | Base |
| Floss\_Cost | Floss Cost | Continuous | Base |
| Reasons\_Not\_to\_Floss | Reasons Not to Floss | Open-ended response on why people don't floss | Rejected |

**Use SAS Enterprise Miner to solve the following questions**

* Create a new diagram and add the data file named **conspref.sas7bdat** to the diagram.
* Make sure that the role of **Response\_id** is set to ID, **Reasons\_Not\_to\_Floss** is set to Rejected, and the roles of all other variables as input.
* Explore the base variables and determine which interval variables should be transformed.
* Apply Log transformations on the following variables: **num\_fillings,** **TP\_Cost** and **Floss\_Cost**
* Apply Exponential transformation on the following variable: **Freq\_teethcleaning**

1. Compare and contrast the original variables versus the transformed variables. **(1 point)**

* Apply *k*-means clustering on the data set.
* **Set** the roles for all the descriptor variable roles (as indicated in the table above) to **No**. This will ensure that you are using the bases only (only transformed or original but not both) for clustering.
* For internal standardization, use **Range**.
* For number of clusters, use **Automatic** as Specification Method, and for Selection Criterion, use **Average** as Clustering Method.
* Examine the cluster results.

1. How many clusters were selected by SAS Enterprise Miner? What are their relative sizes? **(0.5 points)**
2. Look at the cluster history and the CCC plot and comment on the selection of the number of clusters by SAS Enterprise Miner. What other possible cluster solutions are being suggested by the cluster history or CCC plot? **(1 point)**
3. Which are the *four* most important base variables in this cluster solution? **(1 point)**
4. Do the cluster centers seem to be separated in the multidimensional space? **(1 point)**

* Add another Cluster node to conduct a k-Means analysis with 6 clusters.

1. How do the important variables compare between the two cluster sets? **(1 point)**
2. Comparing the information from questions 1-5 above to the new cluster set. Which cluster set would you recommend to the dental association (3 clusters or 6 clusters)? **(1 point)**

* Profile your selected cluster solution via Segment Profile node using the raw bases only
* Make sure that you use a metadata node to set the roles of the transformed bases to rejected and the roles of raw bases to Input and raw bases Hide as **No**

1. Which is the most important variable for **segment 2**? Which is the most important variable for **segment 3**? (**0.5 point**)

* Profile the cluster solution via Segment Profile node using the descriptors only
* Set minimum worth to 0.001

1. Which are the top-2 important variables for **segment 2**? **(0.5 point)**
2. Provide a summary of the cluster profiles with the bases and descriptors and compare the results. **(1.5 points)**
3. What suggestions do you have for the dental non-profit group in relation to the marketing campaign? **(1 point)**

Deliverables:

As you complete the exercise, create a short report in Microsoft Word and in this report answer the questions in the exercise description. Make sure you print your name, section number, student ID# on the report (front page as well as a running header/footer) and turn-in the report as communicated by your instructor.