Monash University: Assessment Cover Sheet

Student name	Wee		Derrick Zhen-Li	
School/Campus			**Student's I.D. number**	31107478
Unit name	ETW3482 - Data mining for business - S1 2022			
Lecturer's name	Dr Lee How Chinh		**Tutor's name**	Ms Jaya Jothi Padiah
Assignment name	Individual Assignment		**Group Assignment: No** **Note, each student must attach a coversheet**	
Lab/Tute Class: Tute 02		**Lab/Tute Time:**		**Word Count:**
Due date: 25-04-2022		**Submit Date: 24/4/2022**		**Extension granted** □

If an extension of work is granted, specify date and provide the signature of the lecturer/tutor. Alternatively, attach an email printout or handwritten and signed notice from your lecturer/tutor verifying an extension has been granted.

Extension granted until (date): ..../......./........... Signature of lecturer/tutor: ....

Late submissions policy	**Days late**	**Penalty applied**
Penalties apply to late submissions and may vary between faculties. Please refer to your faculty's late assessment policy for details.		

Patient/client confidentiality: Where a patient/client case study is undertaken a signed Consent Form must be obtained.

Intentional plagiarism or collusion amounts to cheating under Part 7 of the Monash University (Council) Regulations

Plagiarism: Plagiarism means to take and use another person's ideas and or manner of expressing them and to pass these off as one's own by failing to give appropriate acknowledgement. This includes material from any source, staff, students or the Internet - published and unpublished works.

Collusion: Collusion means unauthorised collaboration on assessable written, oral or practical work with another person. Where there are reasonable grounds for believing that intentional plagiarism or collusion has occurred, this will be reported to the Associate Dean (Education) or nominee, who may disallow the work concerned by prohibiting assessment or refer the matter to the Faculty Discipline Panel for a hearing.

Student Statement:

- I have read the university's Student Academic Integrity Policy and Procedures
- I understand the consequences of engaging in plagiarism and collusion as described in Part 7 of the Monash University (Council) Regulations (academic misconduct).
- I have taken proper care to safeguard this work and made all reasonable efforts to ensure it could not be copied.
- No part of this assignment has been previously submitted as part of another unit/course.
- I acknowledge and agree that the assessor of this assignment may, for the purposes of assessment, reproduce the assignment and:
  i. provide it to another member of faculty and any external marker; and/or
  ii. submit to a text matching/originality checking software; and/or
  iii. submit it to a text matching/originality checking software which may then retain a copy of the assignment on its database for the purpose of future plagiarism checking.
- I certify that I have not plagiarised the work of others or participated in unauthorised collaboration or otherwise breached the academic integrity requirements in the Student Academic Integrity Policy.

Date: .24./..04./..2022... Signature: *Derrick Wee Zhen-Li**

Privacy Statement:

For information about how the University deals with your personal information go to http://privacy.monash.edu.au/guidelines/collection-personal-information.html#enrol

Stage 1: Appropriate Data

This dataset consists of customer information from an automobile company. It has 8,068 observations with 9 unique variables. The company would like to understand its targeted customer segments to market its existing car models.

ID: Unique Customer ID

Gender: Customer's gender.

Ever_Married: Has the customer been married before. Answers consist of Yes and No

Age: Age of customers.

Graduated: Is the customer a graduate holder? Answers consist of Yes and No.

Profession: Customer's occupation of work.

Work_Experience: Number of working experience in years.

Spending_Score: Customer spending capability on vehicle purchase. Ranges from low, medium, high

Family_Size: Number of family members living in a household for the customer (including the customer).

Types	**Count**	**Variables**
Numeric	1	Age
Categorical	8	Gender, Ever_Married, Graduated, Profession, Work_Experience, Spending_Score, Family_Size

Stage 2: Defining a Business Problem and Goal

An automotive company wants to develop sound marketing strategies to drive its business growth. They need to understand customer needs to create new marketing campaigns to promote their existing products of vehicles.

This would require segmentation and profiling of customers using clustering analysis. The data gathered will be from their existing database consisting of demographic information of their consumers. As a starting speculation, relevant variables such as age, spending score, marital status, and family size might be useful to identify key segments for the company answer to these questions of their concern:

- Whom should our advertisements appeal to? Younger or older audiences?
- Where should the advertising budget be allocated? More for luxurious or budget cars?

- What should be the company’s message in their advertisements based on the types of consumers found?

The identified business problems regarding how the company should handle its marketing strategies will be answered in this report further on. In the end, this report has 3 goals to meet mainly:

The goals:

- Identify key customer segments to build customer loyalty and brand recognition.
- Develop a robust marketing engagement plan to increase sales
- Understand customers’ needs and adapt to the fast-changing marketplace

Stage 3: Selecting Inputs

Appendix 1 shows the selection of inputs that would be of interest in conducting k-means clustering. I have selected 4 inputs and rejected 5 variables.

<u>Inputs</u>

In relation to the 4 inputs, the variable “Age” was selected as it was an important feature to find which target segments should the advertisement appeal to more towards the younger or older audience. This will depend on the count for each segment that would arise in the results later. The level of measurement chosen for “Age” is an interval scale as it contains numeric scores.

The “Ever_Married” variable was selected as an input because it was important to know if the segment will be made up of parents, couples, or singles. This input will help identify whether the brand promotion for the company in all their advertisements should be family-oriented or for the single and accomplished style. A binary level of measurement was chosen as there are only two values Yes or No.

The “Spending_score” variable was chosen because it will help determine the spending behaviors of customers. The scores included are low, average, and high spenders. These indications will help me predict whether these customers are more likely inclined to buy luxurious or budget cars from the company. A nominal level of measurement was chosen for this.

Lastly, the “Family_Size” variable was selected to estimate how likely the person would buy family-centered cars with 8-seaters or ordinary cars with 5 seats. This variable will complement the Ever_Married variable to understand customer behaviour. An ordinal level of measurement was chosen because there was a clear ranking between each value. Family sizes in a household of 1-3 are small while 3-5 are medium and 5-9 are large.

Rejected

The first rejected variable was “ID” as it will not provide any statistical help in predicting customers.

“Gender” was excluded as an input because the advertisements should appeal to both male and female audiences and not exclusively. Therefore, when interpreting the segments, it caters to people in general.

“Profession” was excluded as there are many nominal levels of it and it is considered useless. It will not contribute to the development of an advertisement as it should be generalized to every people regardless of the type of job they do.

“Graduate” was an unhelpful indicator as well as it will not provide any meaning to consumers. The advertisement should appeal to people regardless of being a graduated or not.

“Work_experience” was also an unhelpful indicator as it does not matter to the development of the advertisement appeal. It should be advertised regardless of work experience.

Dealing with missing values

Utilizing the StatExplore node, SAS Enterprise Miner identified missing values in some of the categorical variable columns (Appendix 2). According to SAS (n.d.), the StatExplore node is a multipurpose tool used to examine distributions and statistics in datasets. This node was particularly useful in showing the variability and the number of missing values found in the raw dataset. Appendix 3 shows no missing values found for the numeric variable of the age of customers in the dataset. This was fortunate as the handling of missing data will solely focus on categorical variables.

The remedy to solve the missing values located in the dataset will be using complete-case analysis which keeps the rows with observations only. The filter node from SAS Enterprise Miner will aid in this treatment. In the filter node train properties (Appendix 4), I have chosen “Rare Values (Count)” as the default filtering method, and I have specified to not keep missing values by saying “No” for “Keep Missing Values” property under “Class Variables”. “Rare Values (Count)” will instruct SAS to drop value counts that have less than the level specified in the Minimum Frequency Cutoff point. The default value is 1 for the Minimum Frequency Cutoff, therefore SAS EM will exclude any variables in rows that are 0 or blank in the dataset.

Running the filter node had a successful output. Appendix 5 showed 463 observations to have been excluded from the dataset. The remaining observations left were 7605 rows which showed approximately 5.7% of missing values were dropped from the original dataset Thus, this was considered an acceptable rate as it had a very low number of missing values that is lower than many educational and psychological studies having 15-20% missingness (Enders, 2003).

Imputation with mode had been considered as an alternative method but it did not provide good segmentation results as there were incomplete input results in each segmentation when trying many different combinations of clusters ranging from 2-8. Thus, I have used the filter function to exclude the 463 observations as I found good segmentation results in the end.

Next, I implemented another StatExplore Node by connecting it to the Filter node to verify that missing values have been excluded. Indeed, there were no missing values found in any of the categorical variables in the new output of 7605 rows of observations (Appendix 6).

As the observations have been excluded, the inputs of Ever_Married had reduced the percentage of variability from 28.6% to 13.04% in variability (Appendix 7). Lower variability is desirable as Lauer, Kleinman, & Reich (2015) discovered that decreases in cluster size variability lead to an increase in statistical power. Therefore, these inputs can better predict information about our customers in cluster analysis and segmentation reports.

Stage 4: Constructing Clusters

Initial stages of finding the number of clusters started by setting the specification method to "Automatic" in the cluster node. Setting it to automatic will allow SAS EM to quickly determine the optimal number of clusters I need. I left "Internal Standardization" to its default method of "Standardization" as I want my variable values to be divided by standard deviations and not be bounded by range. Furthermore, I changed the clustering method from its default "Ward" to "Centroid" to get the distance between two clusters using the (squared) Euclidean distance between their centroids or means. Ordinal and nominal encoding was set to their defaults of "Rank" and "GLM" as changing them to other methods of encoding didn't help in getting good segmentation results. I concurrently ran my segmentation profiling node with the clustering node to check whether the clusters entered were a good fit for my final segmentation profiling. As I ran the cluster node with the above specifications, I did not end up with complete results in my segmentation profiling. This became an issue and I looked at the selection criterion property panel under the clusters node to start solving this.

Under the "Selection Criterion" properties, there is a customizable section of a number of clusters instead of using user-specify. The reason I did not use user-specify was that I wanted the Cubic Clustering Criterion (CCC) plot to appear in my summary statistics for a justification for the number of selected clusters, however, this was only possible if the specification method was set to "Automatic". Therefore, I had set "Preliminary Maximum" which is the number of clusters used in the preliminary training pass from 50 being its default to 15 because realistically I do not need many clusters to compare. Additionally, I changed the "Final Maximum" that specifies the maximum number of clusters acceptable for the final solution to have the value of 10 instead of its default 20 due to my expectation of clusters being in the range of 3-5. The final configuration of the cluster node panel is indicated in Appendix 8.

The output was successful (Appendix 9), SAS EM produced 4 clusters for me. According to the output, the candidates for an optimum number of clusters were either 4,8 or 10 scorings 90.711, 117.653, and 118.453 respectively on the cubic clustering criterion (Appendix 10). Appendix 11 shows the CCC plot and a modified line graph of the CCC plot to track the peaks of the clusters. Generally, when deciding on the optimal number of clusters, SAS EM will choose a point where the CCC value drops as it increases the number of clusters. The graph in Appendix 11 shows that 4 was one of the optimal numbers of clusters because, at cluster 5, the CCC value started to drop. This could be said the same with the 8^{th} and 10^{th} clusters chosen as another optimal value because, at the 9^{th} and 11^{th} clusters, the CCC value dropped. Conclusively, I have chosen 4 clusters as this was generated fairly by SAS EM based on the CCC plot.

The 4 clusters have potentially 4 unique types of consumers that would be of use to the 3 problems listed by the company. The 4 clusters could be segregated into different types of adults ranging from young, middle-aged, late middle-age, and late adulthood as the age range was found to be between 18 as the minimum and 89 as the maximum. The classification of adults had been based on the ordering made by Medley (1980) in his study of life satisfaction across four stages of adult life. They could also range from being single or married with large, medium, or small families having either low, average, or high spending capability. Additionally, based on their spending capability, these 4 segments could be classified as being in the upper, middle, or lower class of society. Classifying segments into different types of adults that correspond to different ages, spending behaviour, and marital status will help determine the generation appeal of the advertisement, understanding which price point of cars to advertise and its brand message.

Stage 5: Segment Profiling

After gathering 4 clusters, I can now begin segmenting them to identify factors that differentiate one segment from another. Utilizing the Segment Profiling node, I have remained most of the features to their default value except for “Number of Midpoints”. In the General properties, I set the “Number of Midpoints” to 16 from its default 8 (Appendix 12). The value 16 will increase the number of bins in the distribution histogram for my interval variable “Age”. This will allow better visualization of the age range when I compare it against each segment as I found its default value of 8 midpoints to be small. I have also utilized the summary statistics of “Profile: Variables” (Appendix 13) to get the mean values of age for my segmentation analysis below.

The output from the segment profile node was successful. I was able to assess 4 types of consumers. Since I have many categorical inputs, I will be interpreting many pie charts where the outer circle represents the segment results, while the inner circle is the original dataset results. There will be 1 histogram that represents age as it is an interval variable. The order of segments had been based on the highest to the lowest count.

<u>Segment 1: (2543 count)</u>

The Perfect Family group: (Appendix 14)

This segment contained many married Middle-Aged Adults in their 40s, with 1-3 kids as family sizes range from 2-5 living in a household which includes themselves and their spouse. The segment included more average spenders (74.6%) than high spenders (25.40%) both having counts of 1897 and 646 respectively which suggest they are employed individuals and are successful in life. The mean ages were 46 on average and they are likely to be considered between the ranges of Early middle age & Late middle-aged adults (37-60). This group could belong to a middle and upper class of society as they have average and high spenders. Overall, this segment contains more middle-class parents who have a medium-size families looking to buy a luxurious family-size car.

<u>Segment 4: (1926 count)</u>

The Young and Aspiring: (Appendix 15)

This segment contained many single young adults in their 20s who are either university/college students or fresh graduates who are just starting out in life. All 1926 observations were low spenders (100%) and 92.5% of them were not married. They had family sizes of 3-6 including themselves which suggests they have many siblings living in the same household. The age ranges from 20-to 33 years old and the mean age was 28 on average. Overall, this segment suggests these consumers are in their early adult life and may be unable to afford a vehicle because they have just begun working and haven't accumulated enough wealth but are saving for one in the future.

<u>Segment 2: (1700 count)</u>

The Seniors group: (Appendix 16)

This segment consists of many married late middle-aged adults in their 60s, having 1 kid or none as they contained family sizes of 1-3 including themselves and their spouse. This segment accounted for 69.5% of low spenders and 29.9% of high spenders both having 1182 and 518 observations respectively. The age ranges from 46-82 years old with a mean age of 62 on average. Overall, this segment implies that these consumers are in their late middle age or late adulthood life and the majority spend within their means to get through the day to support their families. However, there are a few who are accomplished with the exception of wanting to purchase an expensive class car. This segment is seen as a mixture of old adults wanting to either buy budget or luxurious cars for style and pleasure as they don't need a family-sized car.

<u>Segment 3: (1436 count)</u>

The Living Independently group: (Appendix 17)

This segment contained many single adults in their 30s with family sizes of 1-2 including themselves living in a household which could be either their sibling, parent, or partner. All 1436 observations were low spenders (100%) and 90.1% were not married individuals. They were aged 24-42 years old and had a mean age of 36 on average. Overall, this segment suggests these consumers are likely to buy budget cars for style and pleasure as well.

<u>Interpreting Pie Charts</u>

Spending_Score:
RED = Low,
YELLOW = Average,
BLUE = High

Ever_Married:
RED = Yes,
BLUE= No

Family_size:
GREEN = 1
BROWN = 2
PURPLE = 3
LIME GREEN = 4
LIGHT BLUE = 5
PINK = 6
DARK BLUE = Others

<u>Histogram</u>

Age:
BLUE Bars = segment results,
RED Bars = original dataset results

Stage 6: Implement Segmentation

"Whom should our advertisements appeal to? Younger or older audiences?"

In relation to the first business problem, the automotive company wants to know the most effective target audience to reach out and the answer is older audiences. The company could reach out to either segments 1, 2 or 3 or both as there were many people aged 37-82 years old. Segment 4 should be treated with caution as there may be students and fresh graduates who cannot afford a vehicle due to the lack of accumulated wealth. Therefore, to avoid foreseeable losses the company should avoid targeting segment 4 and should concentrate on older audiences as these people possess higher spending capabilities even though some may be low spenders they would have at least earned some accumulated wealth over time. Conclusively, older audiences dominate the number of potential customers so the company's advertisements should focus on this group to increase their chances of sales and brand awareness.

"Where should the advertising budget be allocated? More for luxurious or budget cars?"

Next, regarding the second business problem, the company wants to know where should the advertising budget be apportioned more into the product lines they own? Their product lines include luxurious and economical class vehicles. Based on the segments found, segments 1 and 2 had a total count of 3061 people who are willing to spend average and high amounts. Conversely, segments 3 and 4 had many people who are willing to spend lower amounts totalling up to 3362 people. Hence, the advertising budget should be allocated approximately 60% to economical class vehicles and 40% to luxurious class vehicles as there is a close-ratio of people likely to purchase either one of these classes.

"What should be the company's message in their advertisements based on the types of consumers found?"

Based on Segment 1 being the highest count of family-oriented consumers, the advertisements could promote safety, comfort, and spacious seating in their vehicles. These consumers are likely looking for comfortable seating and extra space to put many possessions relating to kids, furniture, luggage, and groceries. This message could be used in their luxurious and budget family cars. The alternative would be to display advertisements of style and pleasure as segments 3 and 2 showed no desire in purchasing a family type of vehicle. These segments had a smaller number of households of 1-3, therefore they would prefer more of an elegant design of vehicles with compact spacing.

In conclusion, I have identified 4 key customer segments that could build customer loyalty and brand recognition for the company. I have also determined their main marketing engagement plan which should be directed towards the older audiences looking to purchase luxurious or budget vehicles with features including safety, comfort, and space for the family-oriented segments or the style and pleasure for the independent adults and seniors group. Additionally, I have provided an initial analysis of customer needs for each segment based on the 4 variables I had chosen.

Stage 7: Limitation and Future Works

The segments generated in SAS were considered small therefore it would produce a small profit turnover if the company chose to specifically target one of the segments. Working with a larger dataset that contains 100,000 observations or above will provide more value to these segments to determine their lucrative potential if the company decided to directly target its marketing plans into one of the segments. Future works would suggest adding more observations from their database to determine how much more profit they could potentially earn.

Consumer analysis in this report might be misinterpreted. The interpretation made in the segmentation results was limited to the 4 variables I have chosen based on spending score, family size, marital status, and age. Many other variables would better study consumer behaviors of buying a vehicle which could include the other 5 variables rejected mainly, gender, profession, graduate, and work experience. Future works would suggest adding more variables to better contextualize consumers' needs and demands through studying the other 5 rejected variables and conducting additional customer needs assessments and satisfaction surveys.

References

Enders, C. K. (2003). Using the Expectation Maximization Algorithm to Estimate Coefficient Alpha for Scales With Item-Level Missing Data. *Psychological Methods*, *8*(3), 322–337. https://doi.org/10.1037/1082-989X.8.3.322

Lauer, S. A., Kleinman, K. P., & Reich, N. G. (2015). The effect of cluster size variability on statistical power in cluster-randomized trials. *PLoS One*, *10*(4), e0119074. https://doi.org/10.1371/journal.pone.0119074

Medley, M. L. (1980). Life Satisfaction across Four Stages of Adult Life. *The International Journal of Aging and Human Development*, *11*(3), 193–209. https://doi.org/10.2190/D4LG-ALJQ-8850-GYDV

Appendixes

(Appendix 1)

Variables - FIMPORT

(none) not Equal to ...

Columns: Label Mining

Name	Role	Level	Report	Order	Drop	Lower Limit	Upper Limit
Age	Input	Interval	No		No	.	.
Ever_Married	Input	Binary	No		No	.	.
Family_Size	Input	Ordinal	No		No	.	.
Gender	Rejected	Binary	No		No	.	.
Graduated	Rejected	Binary	No		No	.	.
ID	Rejected	Nominal	No		No	.	.
Profession	Rejected	Nominal	No		No	.	.
Spending_Score	Input	Ordinal	No		No	.	.
Work_Experienc	Rejected	Ordinal	No		No	.	.

(Appendix 2): Missing values for categorical

Output

```
Class Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN

                                        Number
Data                                      of                              Mode                     Mode2
Role     Variable Name       Role       Levels    Missing    Mode       Percentage    Mode2         Percentage

TRAIN    Ever_Married        INPUT         3         140     Yes          57.55       No              40.72
TRAIN    Family_Size         INPUT        10         335     2            29.62       3               18.55
TRAIN    Gender              INPUT         2           0     Male         54.75       Female          45.25
TRAIN    Graduated           INPUT         3          78     Yes          61.58       No              37.46
TRAIN    Profession          INPUT        10         124     Artist       31.18       Healthcare      16.51
TRAIN    Spending_Score      INPUT         3           0     Low          60.46       Average         24.47
TRAIN    Work_Experience     INPUT        16         829     1            29.18       0               28.73
```

(Appendix 3): Missing values for numeric

```
Interval Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN
```

Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
Age	INPUT	43.46691	16.7117	8068	0	18	40	89	0.696021	-0.14545

(Appendix 4): Filter Node Train Property

Property	Value
General	
Node ID	Filter
Imported Data	...
Exported Data	...
Notes	...
Train	
Export Table	Filtered
Tables to Filter	Training Data
Distribution Data Sets	Yes
Class Variables	
Class Variables	...
Default Filtering Method	Rare Values (Count)
Keep Missing Values	No
Normalized Values	Yes
Minimum Frequency Cutoff	1
Minimum Cutoff for Percent	0.01
Maximum Number of Levels	25
Interval Variables	
Interval Variables	...
Default Filtering Method	None

(Appendix 5): Filter Node output

Output

```
Number Of Observations

Data
Role      Filtered    Excluded    DATA

TRAIN       7605         463      8068



Statistics for Original and FILTERED Data
(maximum 500 observations printed)

Data Role=TRAIN Variable=Age

Statistics               Original    Filtered

Non Missing               8068.00     7605.00
Missing                      0.00        0.00
Minimum                     18.00       18.00
Maximum                     89.00       89.00
Mean                        43.47       43.53
Standard Deviation          16.71       16.63
Skewness                     0.70        0.69
Kurtosis                    -0.15       -0.14


*------------------------------------------------------------*
* Report Output
*------------------------------------------------------------*
```

(Appendix 6): 2nd StatExplore connecting Filter Node

Results - Node: StatExplore (3) Diagram: A1 v.2

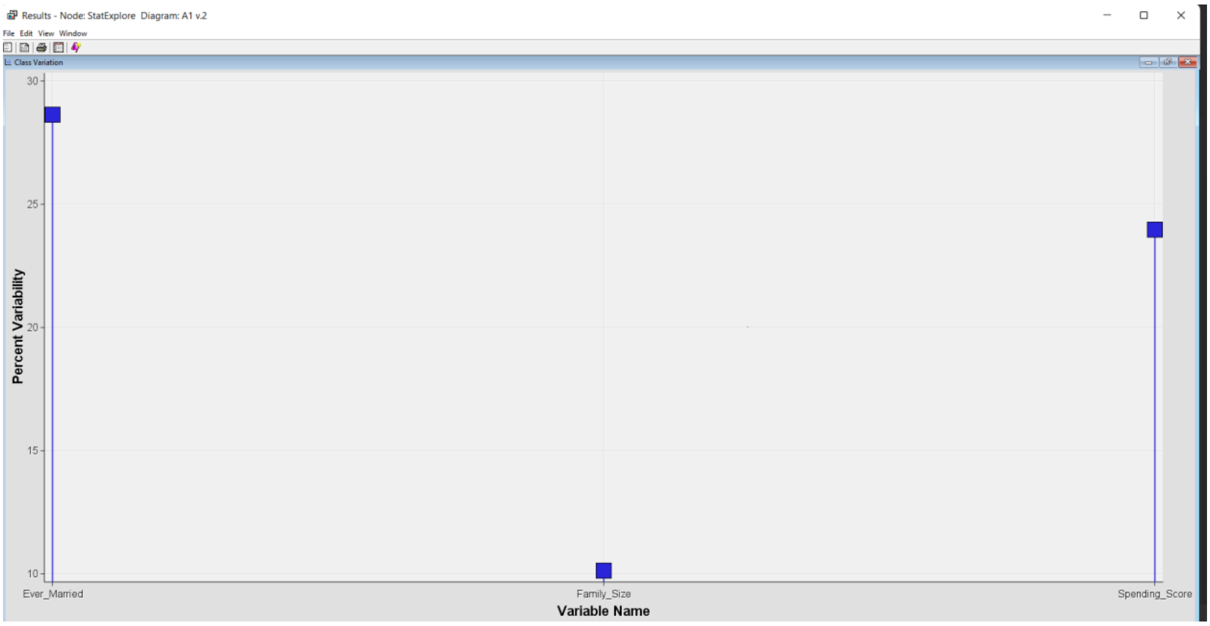
File Edit View Window

Output

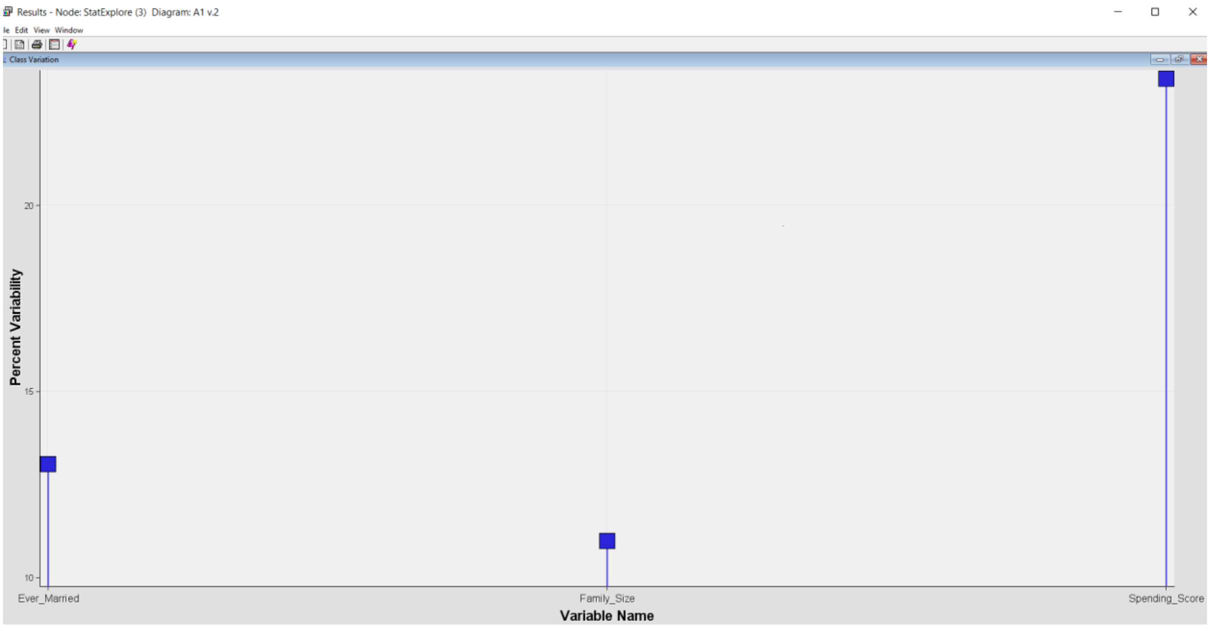
```
Time:                08:40:21
*------------------------------------------------------------*
* Training Output
*------------------------------------------------------------*




Variable Summary

         Measurement    Frequency
Role        Level         Count

ID        NOMINAL           1
INPUT     BINARY            1
INPUT     INTERVAL          1
INPUT     ORDINAL           2



Variable Levels Summary
(maximum 500 observations printed)

                    Frequency
Variable    Role      Count

   ID        ID        7605



Class Variable Summary Statistics
(maximum 500 observations printed)

Data Role=TRAIN

                                    Number
Data                                  of                               Mode                  Mode2
Role     Variable Name      Role    Levels    Missing    Mode    Percentage    Mode2       Percentage

TRAIN    Ever_Married       INPUT      2         0       Yes        59.22      No            40.78
TRAIN    Family_Size        INPUT      9         0       2          31.06      3             19.25
TRAIN    Spending_Score     INPUT      3         0       Low        59.76      Average       25.06



Interval Variable Summary Statistics
```

(Appendix 7)

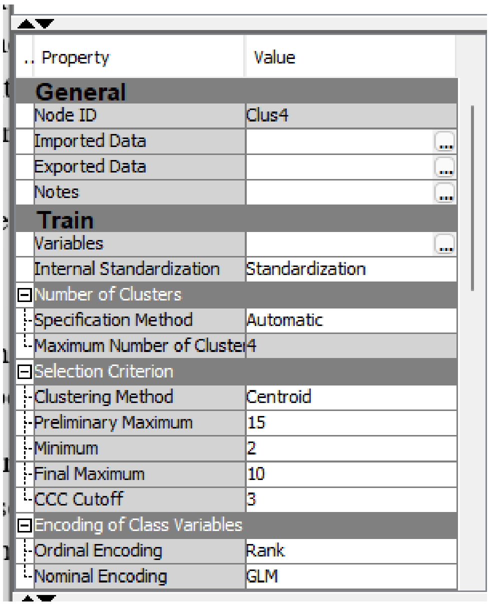
Before



After

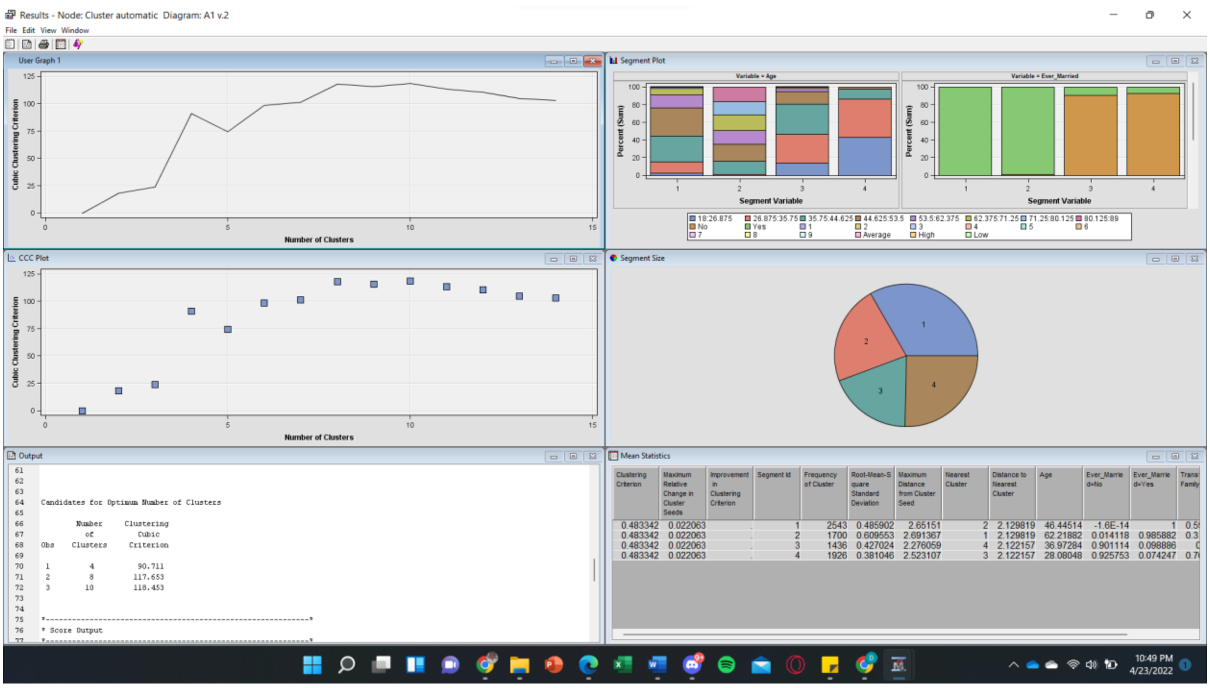


(Appendix 8): Cluster Node



Property	Value
General	
Node ID	Clus4
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
Internal Standardization	Standardization
Number of Clusters	
Specification Method	Automatic
Maximum Number of Cluster	4
Selection Criterion	
Clustering Method	Centroid
Preliminary Maximum	15
Minimum	2
Final Maximum	10
CCC Cutoff	3
Encoding of Class Variables	
Ordinal Encoding	Rank
Nominal Encoding	GLM

(Appendix 9): Clustering output



(Appendix 10): Optimal number of clusters

```


Candidates for Optimum Number of Clusters

        Number     Clustering
          of         Cubic
Obs    Clusters    Criterion

 1         4          90.711
 2         8         117.653
 3        10         118.453


*------------------------------------------------------------*
* Score Output
*------------------------------------------------------------*


*------------------------------------------------------------*
* Report Output
*------------------------------------------------------------*


```

(Appendix 11): Cubic Clustering Criterion

(Appendix 12): Segment Profiling Node

Property	Value
Node ID	Prof5
Imported Data	...
Exported Data	...
Notes	...
Train	
Variables	...
General	
Number of Midpoints	16
Profile All	No
Cutoff Percentage	95
Input Variables	
Number of Inputs	10
Minimum Worth	0.01
Maximum Depth	1
Print Worth Statistics	Yes
Target Variables	
Analysis Role	None
Report Variables	
Use Report Variables	Yes
Number of Report Variables	10

(Appendix 13): Summary statistic of profile variables

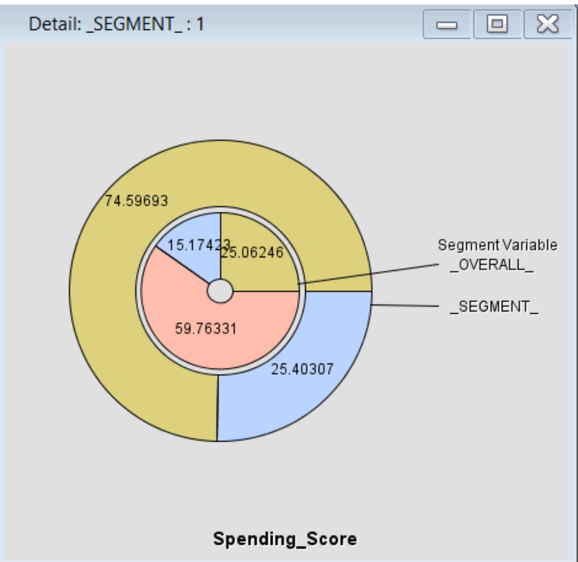
Results - Node: (PICKED) Segment Profile (5) Diagram: A1 v.2

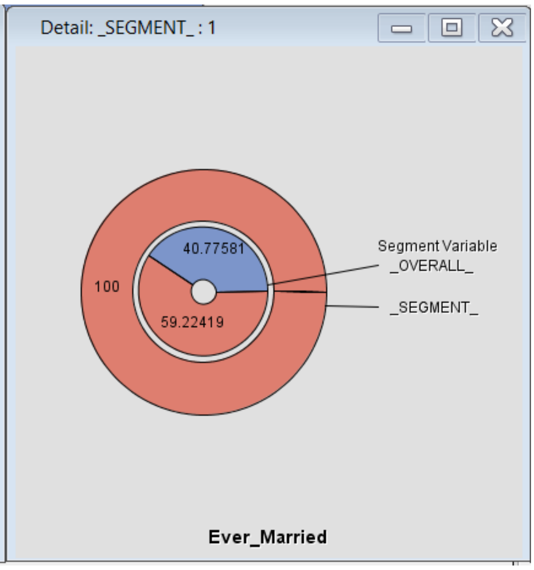
File Edit View Window

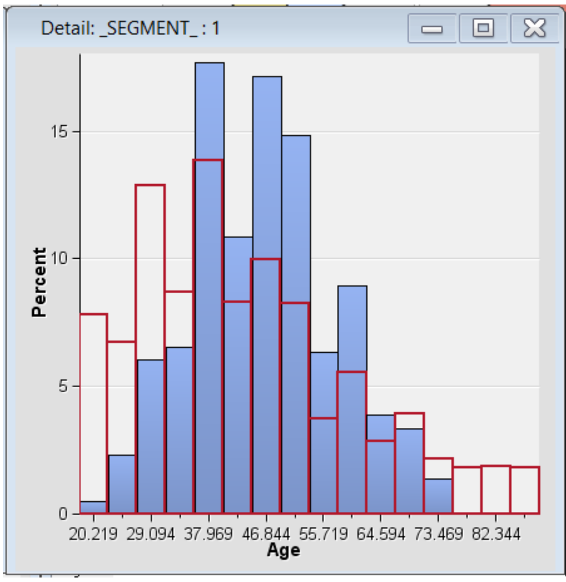
Profile Variables

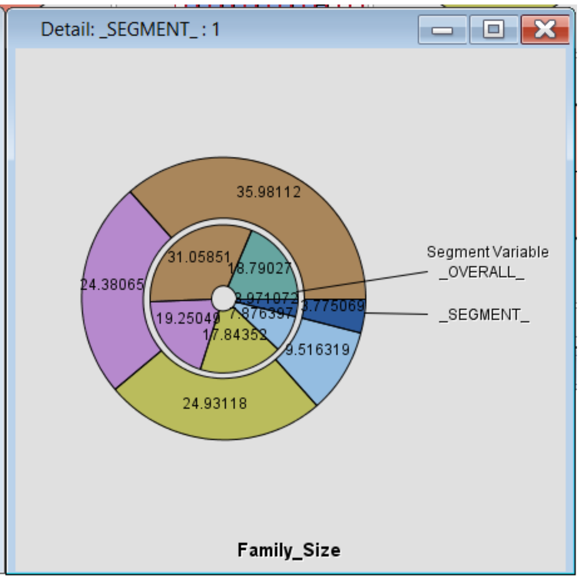
Type	Segment Variable	Segment Value	Variable	Rank	Worth	Label	Number of Levels	Missing	Minimum	Maximum	Mean	Standard Deviation	Skewness	Kurtosis
I	OVERA...	EMWS8....	Age	.	.	Age	.	0	18	89	43.53162	16.63427	0.69473	-0.13708
C	OVERA...	EMWS8....	Ever Mar...	.	.	Ever Mar...	2	.	.	.	.	.	.	.
C	OVERA...	EMWS8....	Family S...	.	.	Family S...	9	.	.	.	.	.	.	.
C	OVERA...	EMWS8....	Spending...	.	.	Spending...	3	.	.	.	.	.	.	.
C	SEGME...	1	Spending...	1	0.368002	Spending...	2	.	.	.	.	.	.	.
C	SEGME...	1	Ever Mar...	2	0.153967	Ever Mar...	1	.	.	.	.	.	.	.
I	SEGME...	1	Age	3	0.083478	Age	.	0	18	89	46.44514	11.13468	0.221437	-0.38428
C	SEGME...	1	Family S...	4	0.05238	Family S...	9	.	.	.	.	.	.	.
I	SEGME...	4	Age	1	0.17349	Age	.	0	18	70	28.08048	7.32491	0.987615	1.693958
C	SEGME...	4	Ever Mar...	2	0.142526	Ever Mar...	2	.	.	.	.	.	.	.
C	SEGME...	4	Family S...	3	0.13227	Family S...	7	.	.	.	.	.	.	.
C	SEGME...	4	Spending...	4	0.086364	Spending...	1	.	.	.	.	.	.	.
I	SEGME...	2	Age	1	0.136959	Age	.	0	35	89	62.21882	15.50778	-0.00688	-1.17376
C	SEGME...	2	Ever Mar...	2	0.064125	Ever Mar...	2	.	.	.	.	.	.	.
C	SEGME...	2	Family S...	3	0.050343	Family S...	8	.	.	.	.	.	.	.
C	SEGME...	2	Spending...	4	0.039862	Spending...	3	.	.	.	.	.	.	.
C	SEGME...	3	Family S...	1	0.107799	Family S...	3	.	.	.	.	.	.	.
C	SEGME...	3	Ever Mar...	2	0.071872	Ever Mar...	2	.	.	.	.	.	.	.
C	SEGME...	3	Spending...	3	0.04801	Spending...	1	.	.	.	.	.	.	.
I	SEGME...	3	Age	4	0.024204	Age	.	0	18	73	36.97284	9.882654	0.770829	0.720772

(Appendix 14): Summary results of Segment 1

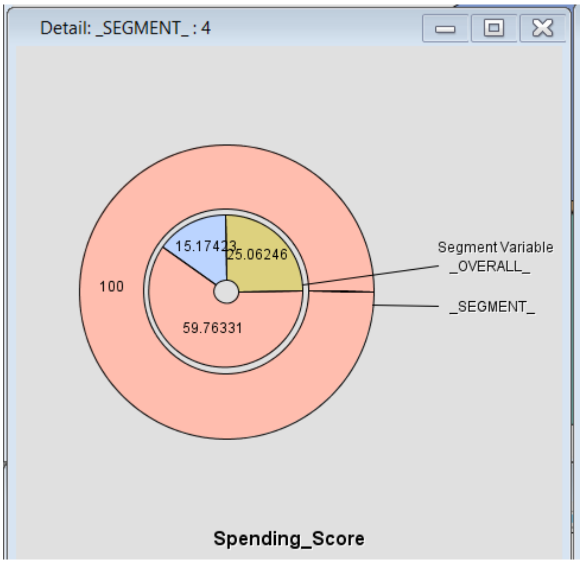


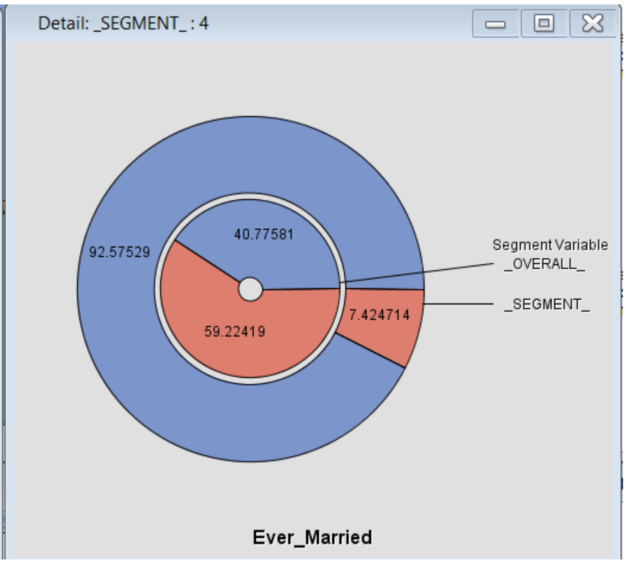


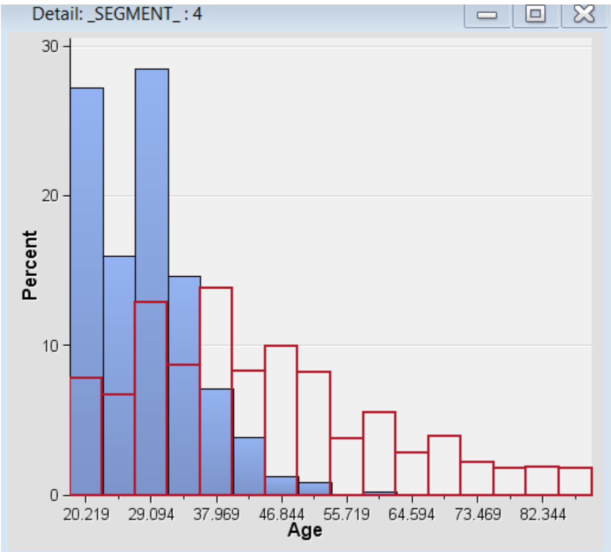


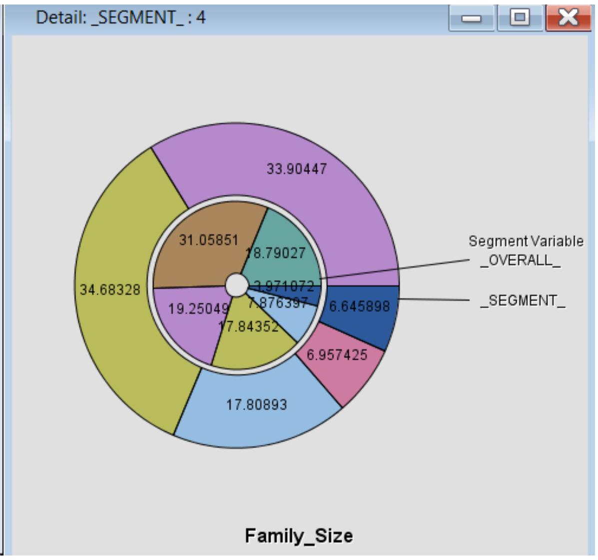


(Appendix 15): Summary results of Segment 4

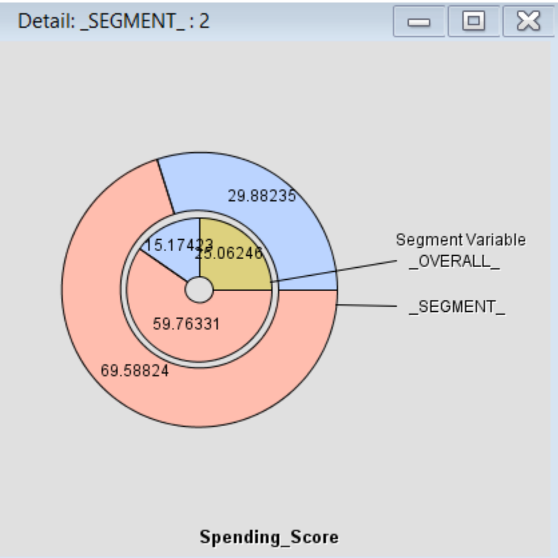


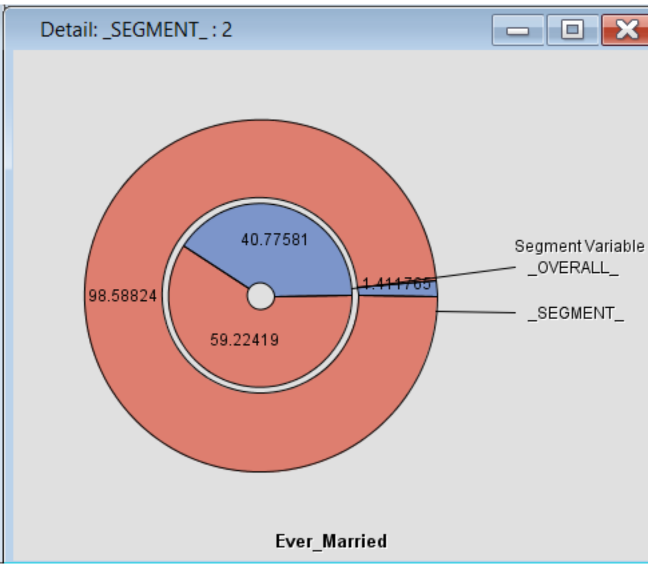


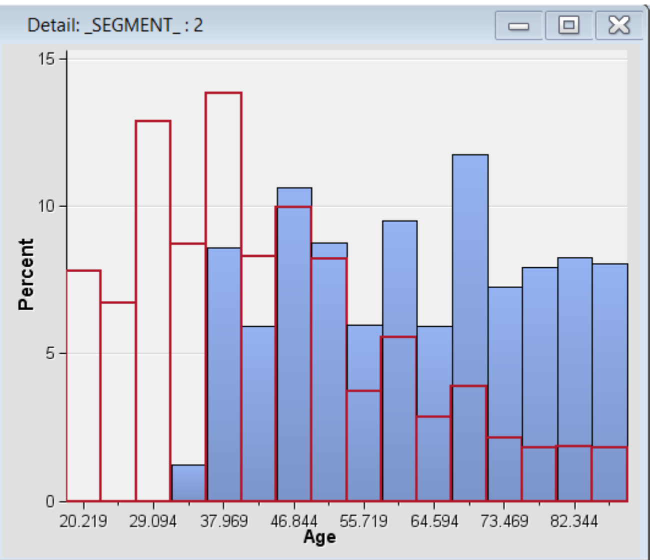


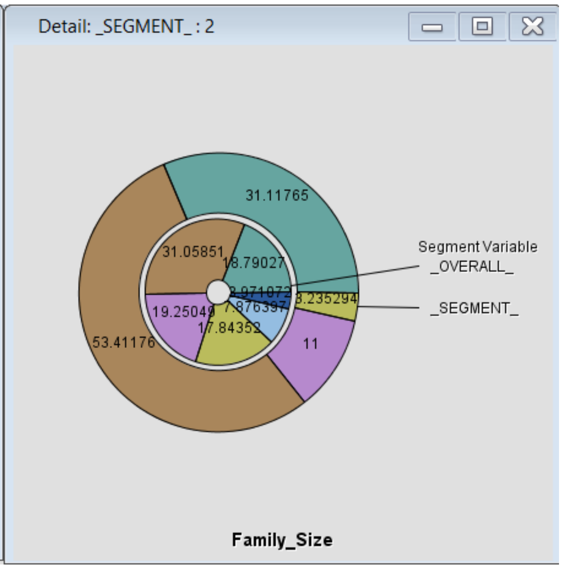


(Appendix 16): Summary statistics of Segment 2









(Appendix 17): Summary statistics of Segment 3