Clustering Bank Customers

What is the marketing business purpose of this assignment?
To segment a bank's customers based on historic usage patterns to identify those who might benefit from new product offerings for a cross-sell/up-sell campaign.

Data Exploration

1. How many rows are in the data set? 100,000
    a. How many Variables? 6
    b. What does each row represent? A record; an individual or household with at least one checking account and at least one transaction on the account during a 3-month study period.

2. **Copy/Paste a Screen shot of the StatExplore node output.**

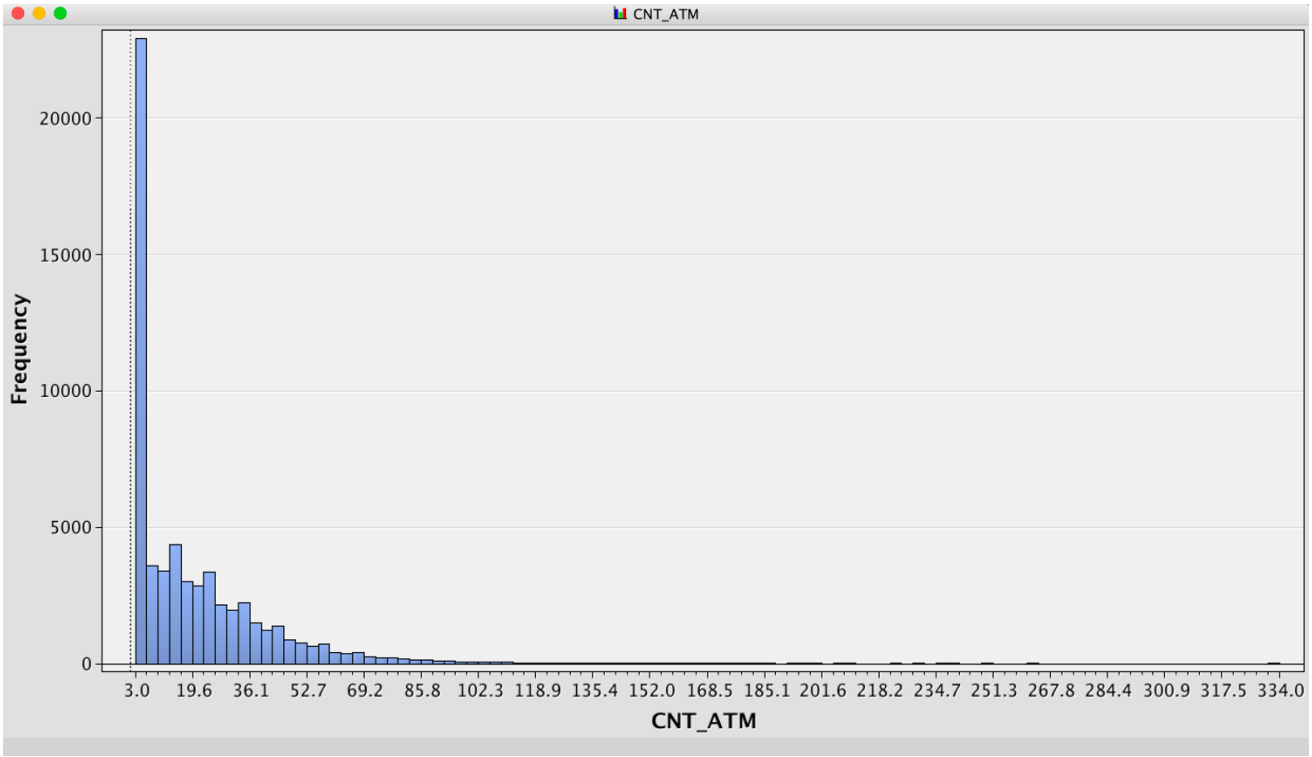
Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
CNT_ATM	INPUT	19.49971	20.8561	100000	0	3	13	628	2.357293	15.33353
CNT_CSC	INPUT	6.68411	12.12856	100000	0	1	2	607	6.236494	108.4633
CNT_POS	INPUT	11.9233	20.73384	100000	0	2	2	345	3.343805	15.43941
CNT_TBM	INPUT	68.13696	101.1542	100000	0	10	52	14934	53.05219	5925.782
CNT_TOT	INPUT	106.2441	113.3704	100000	0	17	89	15225	39.2061	3976.09

3. For CNT_ATM variable:
    a. What is the min and max for customer ATM use over the 3-month period? MIN=3; MAX=628
    b. What does the skew measure indicate? A high, positive level of skew (2.3573); there may be outliers in the data set
    c. What is the best value for the measure of central tendency of these bank customers in terms of ATM use? The median (13)

4. For CNT_TBM variable:
    a. What is the min and max for customer use of traditional bank methods over 3-month period? MIN=10; MAX=14,934
        1) What is the mean number of TBM use of customers? 68.13696
        2) Is the mean a reasonable value for the center of the data for TBM? Why or why not? No, since the data is highly skewed. The median is a better measure of central tendency when the data is skewed.
    b. What is the effect of skew on the mean for TBM use? There may be outliers in the data set that will increase the value of the mean. TBM use has the highest level of skew (53.05219)
    c. In statistics, how is the range calculated? Max – Min
        1) What is the range for CNT_TBM? 14,924

5. Describe the shape of the distributions (histograms) of the variables. They are all highly, positively skewed.

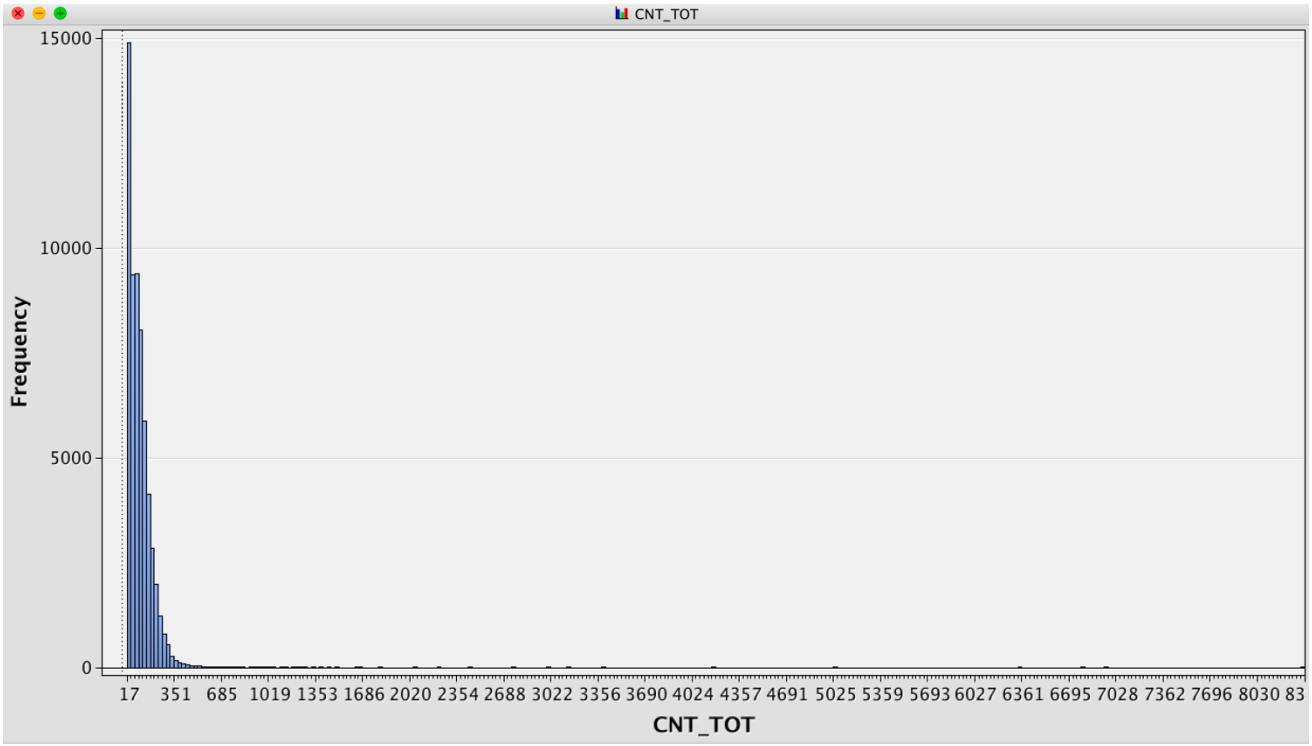
6. Re-bin CNT _ATM to show 100 bars.
    a. **Copy/Paste** the histogram.



b. What is the default bin count? 10
c. What is the purpose of re-binning the histogram? To get a more in-depth view of the records in the histogram, which may show us outliers that we otherwise couldn't see with 10 bins
d. Interpret the tallest bar. There are 22,937 records between 3 and 6.31

7. Re-bin CNT_POS to show 100 bars.
   a. Do customers using the ATM the fewest times also use the bank's debit card (POS) the fewest times? (use interactivity function of histograms) Yes

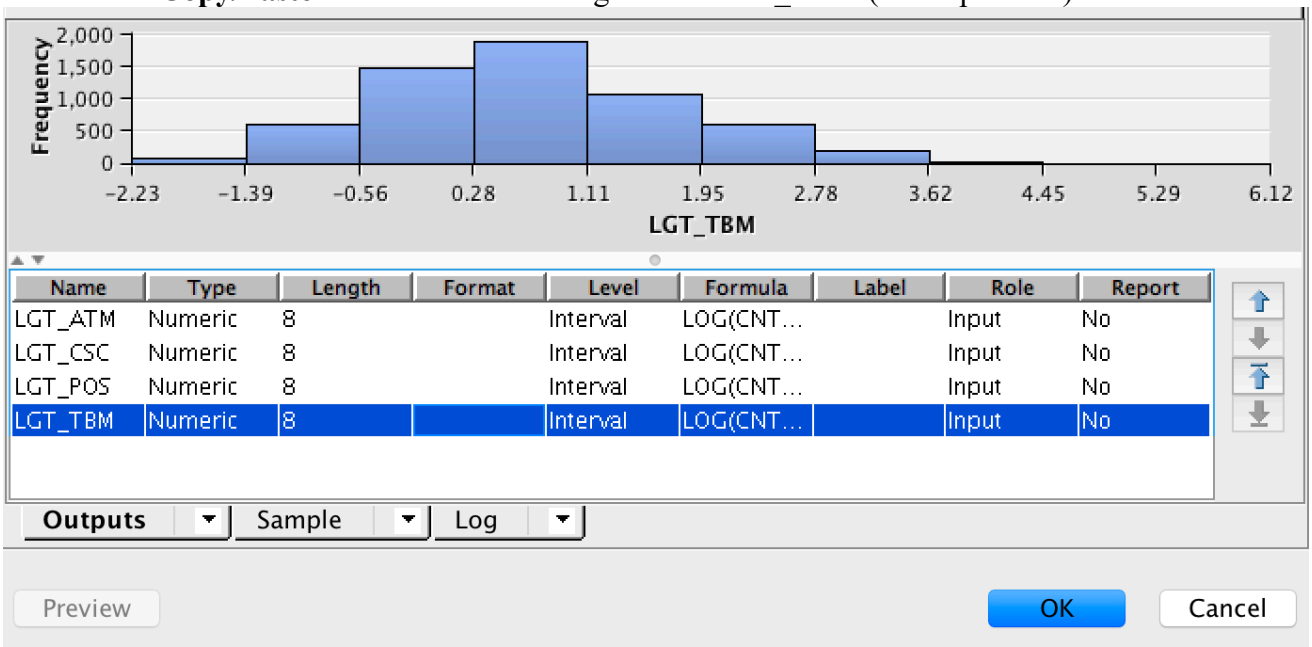
8. Re-bin CNT_TOT to show 300 bars.
   a. **Copy/Paste** the histogram.



b. Interpret the tallest bar. There are 14,833 records between 17 and 44.82333.

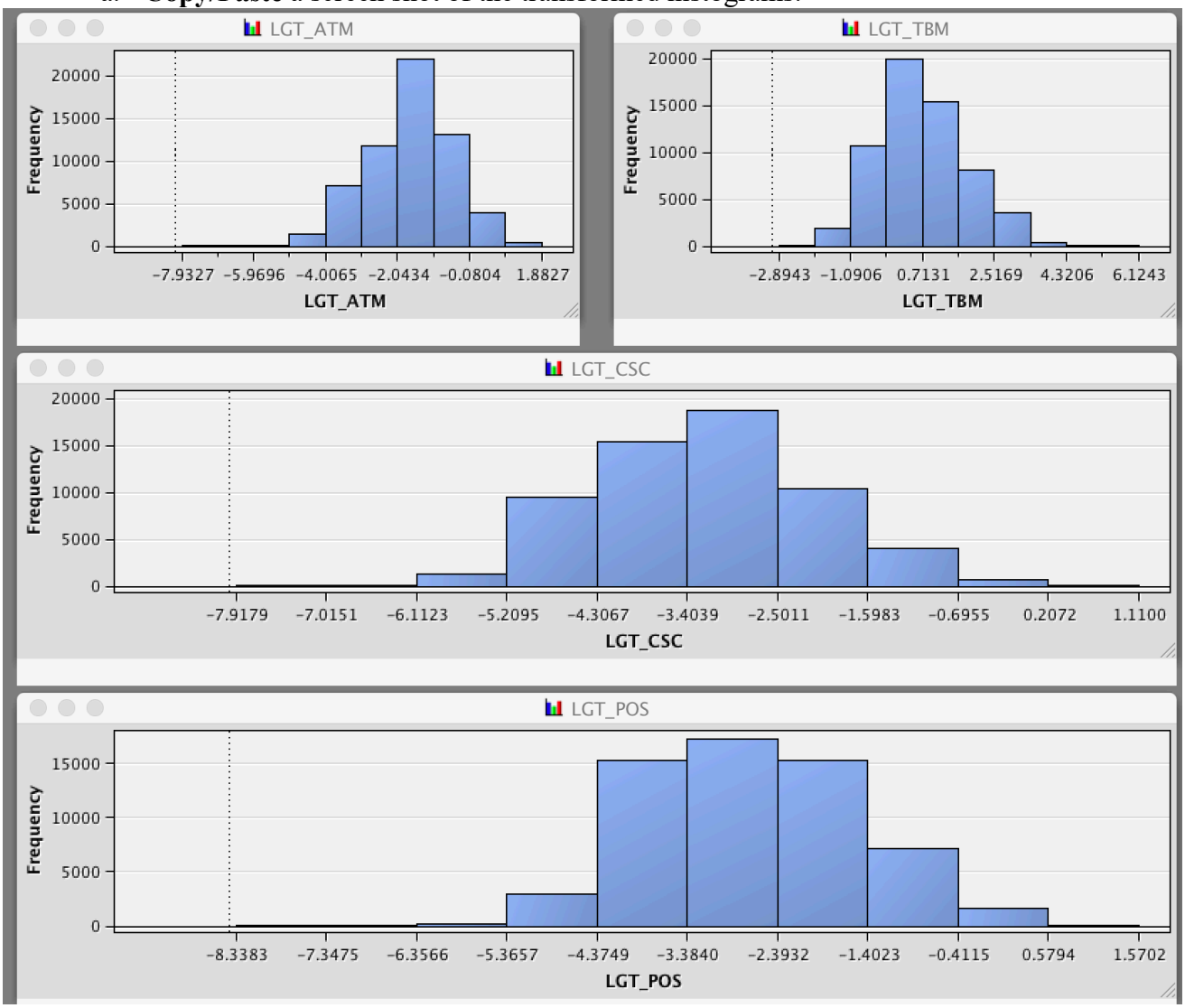
9. Why will we transform the data? It would be difficult to develop meaningful segments from such highly skewed inputs. Instead of focusing on the transaction counts, it was decided to develop segments based on the relative proportions of transactions across the four categories. This required a transformation of the raw data.

10. You will create **4 LOG formulas**. One each for LGT_ATM, LGT_CSC, LGT_POS, and LGT_TBM
    a. **Copy/Paste** the transformed histogram for LGT_TBM (select preview)



11. Run the Transform Node, view the results.
    a. What is the skew for the Log transformed ATM variable? -0.26459
    b. What is the original skew for the TBM distribution? 53.05219
        1) What is the skew after log transformation? 0.350694

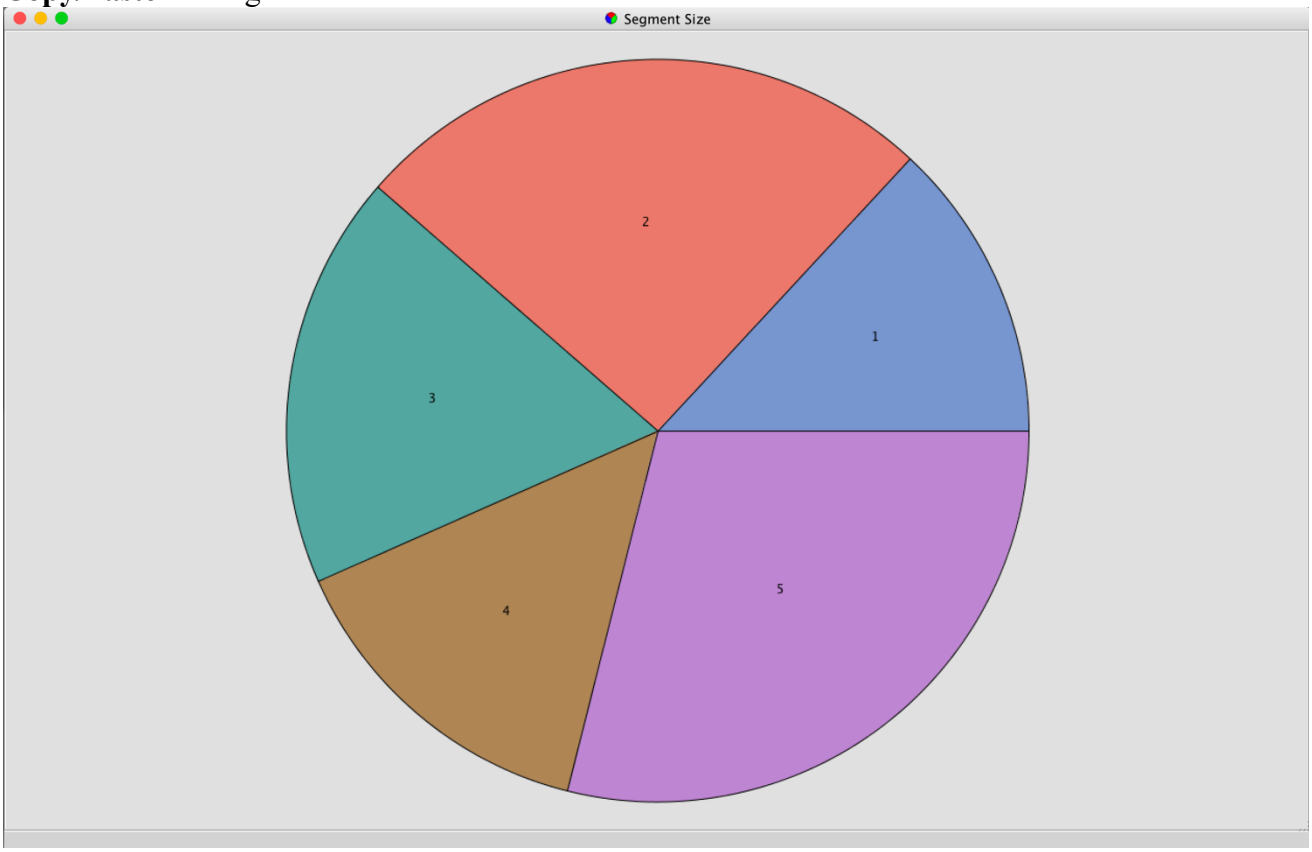
12. Add the Cluster Node to the diagram, **right click** on it, select the 4 transformed variables and click Explore.
    a. **Copy/Paste** a screen shot of the transformed histograms.



    b. Describe the new distributions. They have more of a normal distribution than before.
        1) What statistic would best represent the center of these distributions? The mean since there is low skew, and the histograms have a normal distribution. The mean is the best measure of central tendency when skew is low.

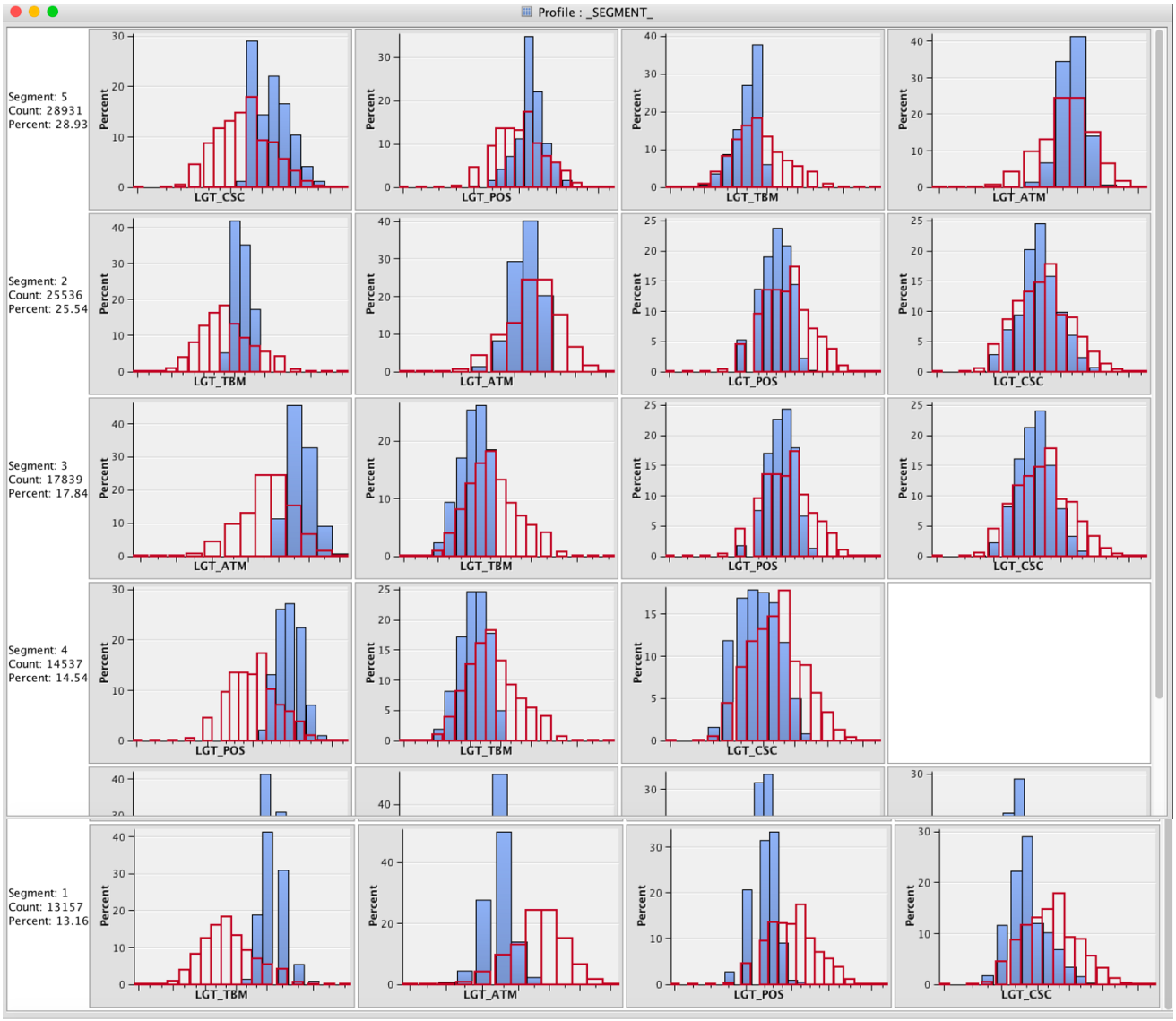
13. Be sure to change the standardization to **None.**
    a. Why are these variables not standardized? The inputs were on the same measurement scale within SAS
    b. Why are 5 clusters a good number for this cluster project? Limiting to 5 clusters with the most amount of records gives us more meaningful results and analyses.

14. **Copy/Paste** the Segment Size Pie Chart.



   a. What is your assessment of the # of records in each cluster for profiling purposes?
   Each of the 5 clusters have an ample amount of records for meaningful profiling.

15. Attach the Segment Profile Node, **(5 'no', 2 'default', 4 'yes', 2 'default') -** run the Profile node.
   a. **Copy/Paste** the profiles after running the Segment Profile Node.



<u>Fill in the following chart</u> with the segment # (cluster) that best matches the given profile name.

Segment #	Customer Profile Name
3	**ATM Customers**
5	**Service Customers (customers with customer service contacts)**
1	**Brick-and-Mortar Customers (customers using bank lobby)**
4	**Cashless Customers (customers using bank debit card)**
2	**Transitional Customers (appear to be transitioning from brick-and-mortar to other bank usage patterns).**

16. It makes most sense to focus on changing the behavior of which set of the above Customers first? Why?
    It would be most monetarily beneficial to change the behavior of brick and mortar customers because banks incur overhead expenses and pay wages to operate brick and mortar locations. The brick and mortar customers should be encouraged to use ATMs and mobile/online banking services.

Fill in the following chart matching a marketing strategy with a bank segment identified in the cluster profile. One segment per strategy and omit the Transitionals because they are already transitioning.

Marketing Strategy	Segment #
Which segment of customers might you want to incentivize to use the ATM card more often?	**3**
To which segment of customers might you promote a reward program for bank debit card use?	**4**
To which segment of customers might you offer a $500 cash prize in a drawing for those who respond to a survey about **why** they use customer services?	**5**
To which segment of customers would you promote a higher checking account interest rate if in lobby banking is limited to twice a month?	**1**