**Cluster Analysis Assignment**

**Data Mining**

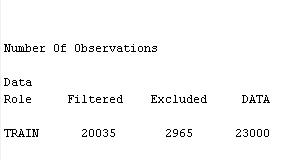
**AMBA**

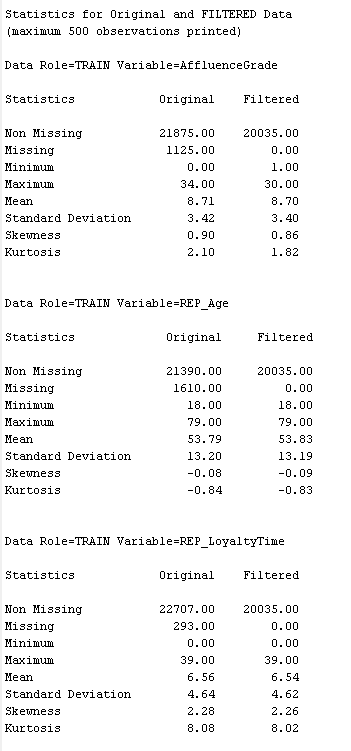
**Mohamed AlaaElDin Mohamed Awad**

**ID: 5910215012**

**Clustering Analysis (Super Market Campaign)**

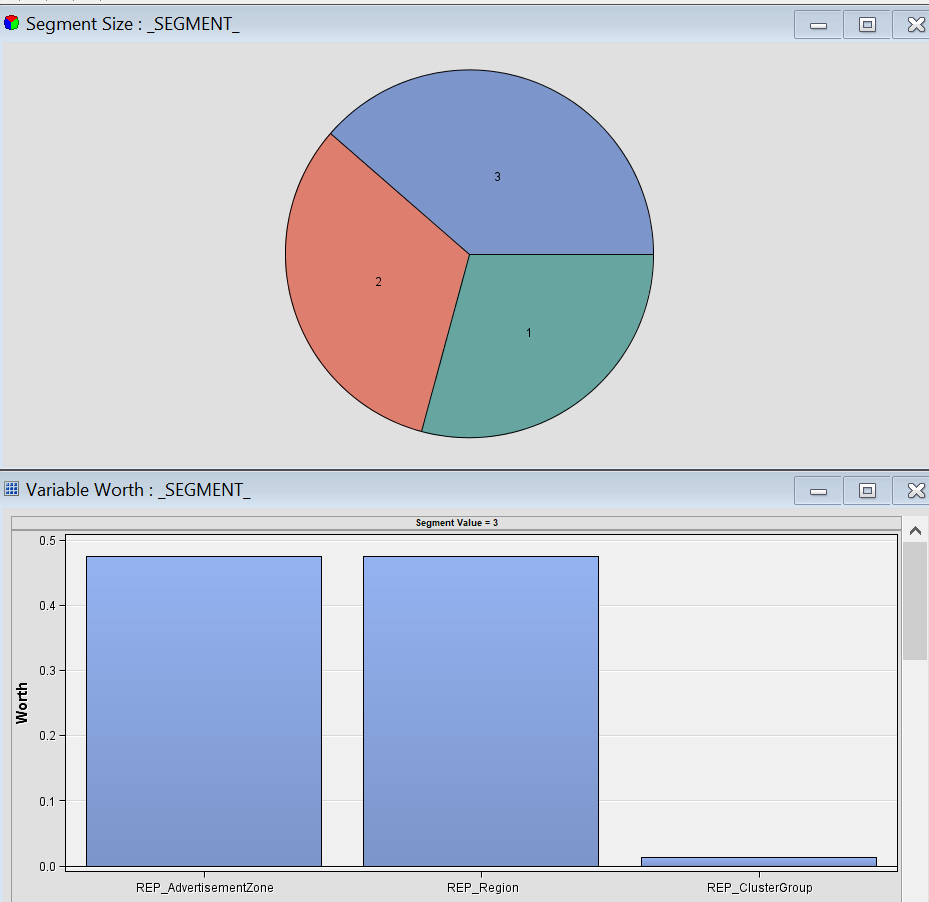
To cluster the customers into segments, it was necessary first to filter the data from any outliers.

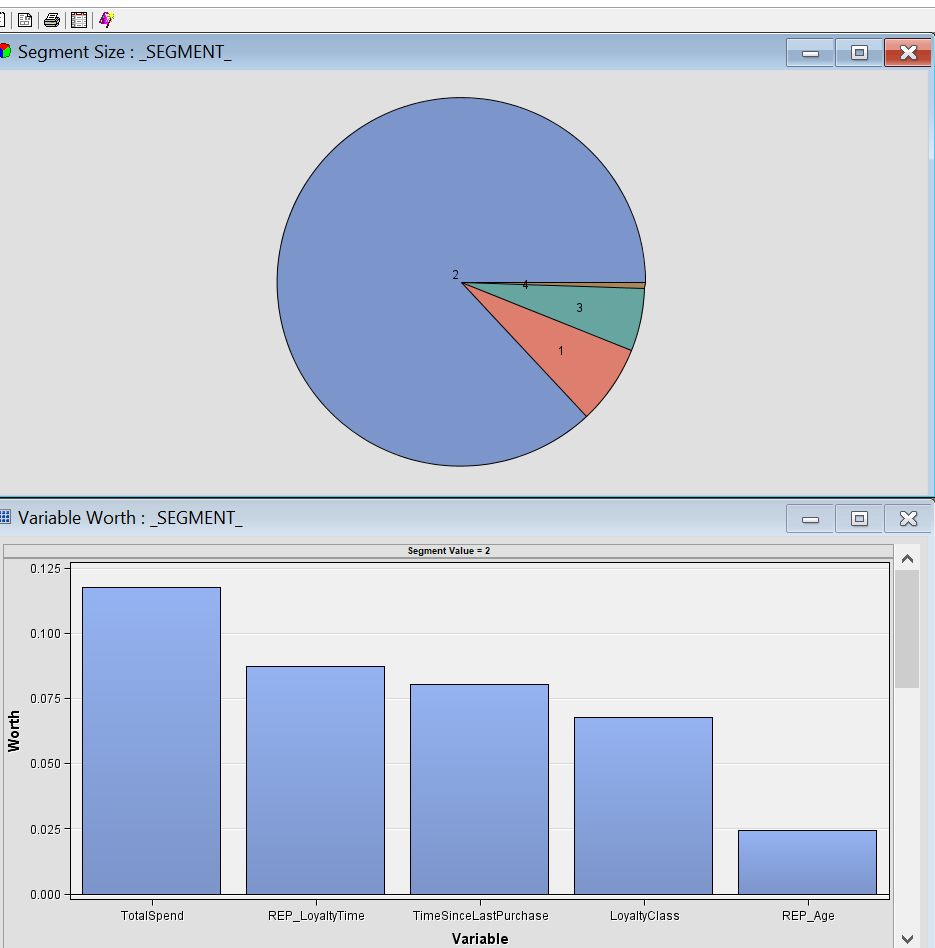


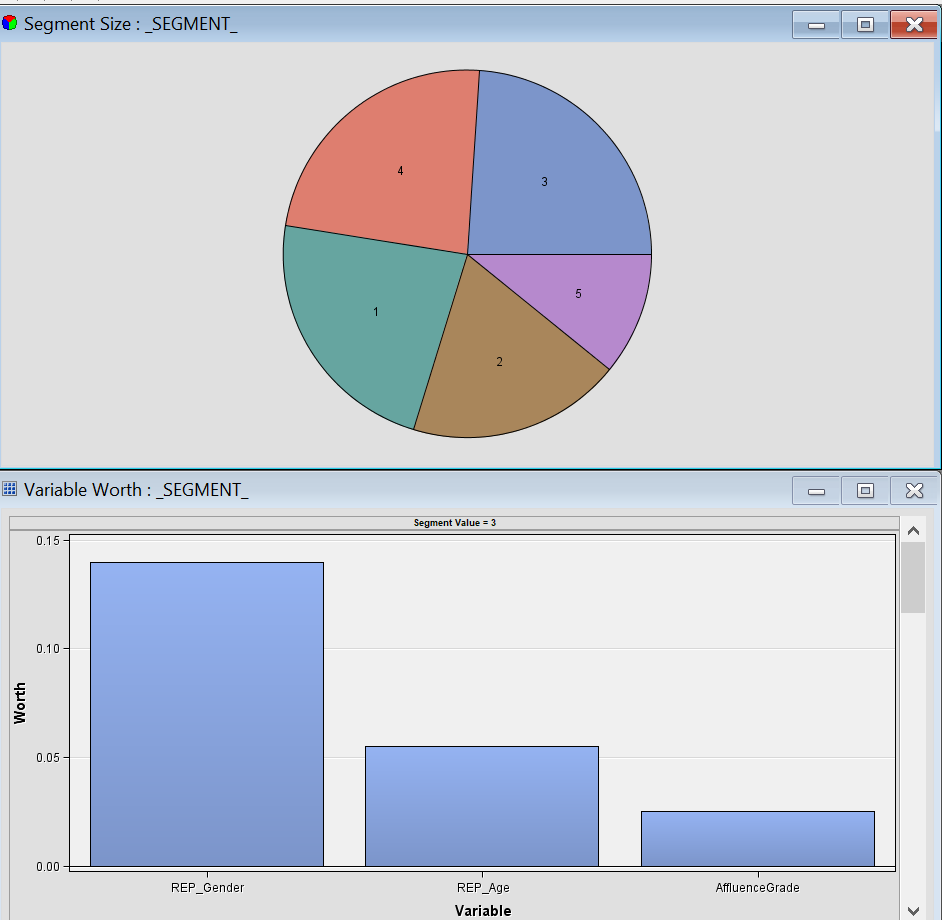


A number of different clusters were tried to find the model that better represents the data and later on building a predictive model with the lowest possible misclassification rate.

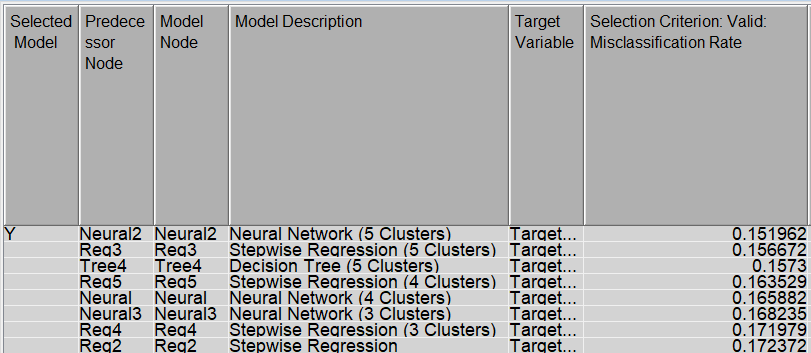
To achieve that, different numbers of clusters were tried: 3 clusters, 4 clusters, and 5 clusters. After that the node “Segment profile” was executed for each of the clusters to have an idea about who are the customers in each cluster.



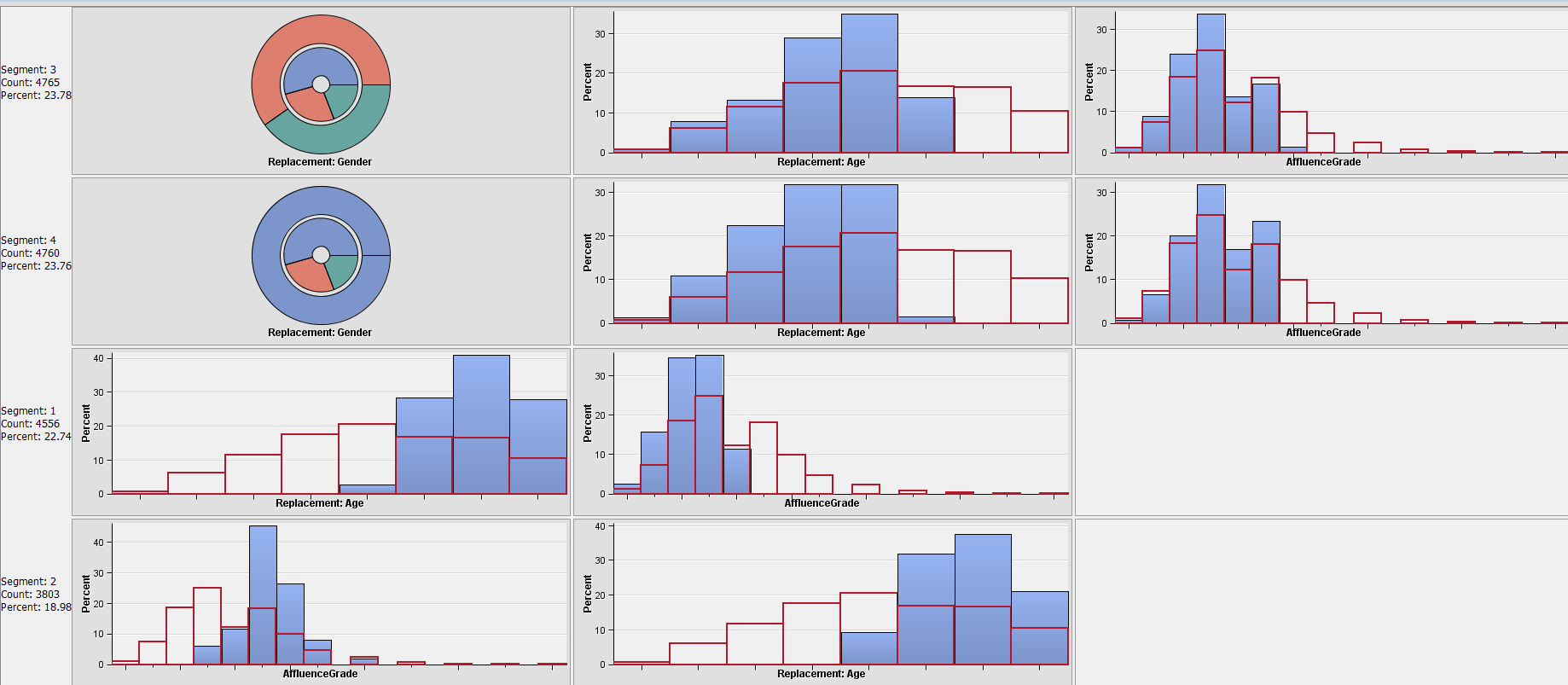


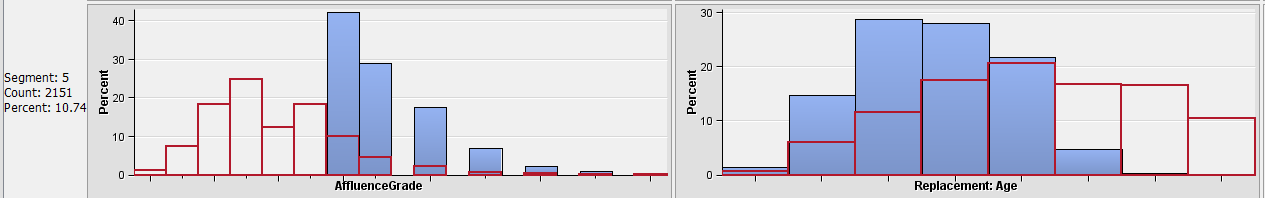


After running predictive modeling for each cluster, it appeared that the model with 5 clusters achieved the lowest misclassification rate of all models.



For neural network model for the 5 cluster segments, many different numbers of hidden units were tried, the lowest misclassification rate came out from the neural network that contains 14 hidden units, the misclassification rate was 0.151962.

The different segments from the 5 clusters model involve the following variables:



Segment 3(Male and working): Contains 4765 customers, represents 23.78 % and can be characterized by Gender(60% are Males), Replacement Age(21-59), Affluence Grade (1.8-12). Those people would need more discounts and sales to become more loyal.

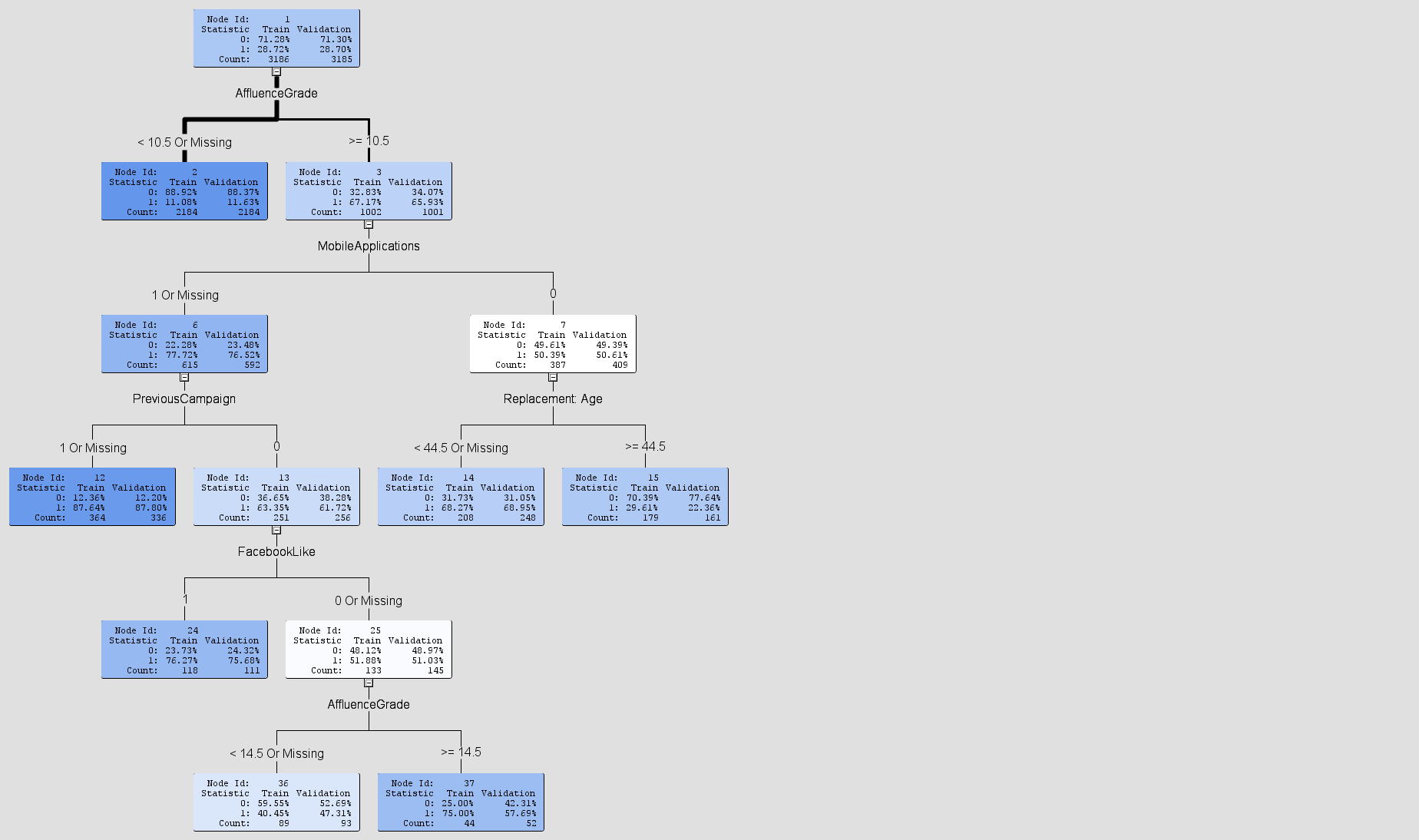
Segment 4(Females and working): Contains 4760 customers, represents 23.76 % and can be characterized by Gender(100% are Females), Replacement Age(21-59), Affluence Grade (1.8-10.6). Those people would need more discounts and sales to become more loyal. Since this group contains females only, they can be targeted by cosmetics products.

Segment 1(Retired and less loyal): Contains 4556 customers, represents 22.74% and can be characterized by age (52-75) and Affluence grade (1.8-8)

Segment 2 (Retired and more loyal): Contains 3803 customers, represents 19 % and can be characterized by Affluence grade (7-17) and age (52-75) . This segment is transitional, that’s, they can become more loyal if you offer them rewards and discounts on certain purchases relative to their age.

Segment 5(Working and the most loyal): ): Contains 2152 customers, represents 10.4 % and can be characterized by Affluence grade (12-26) and age (21-59). Since this group contains the most loyal customers, those customers should be retained through various choices such as offering them rewards on their special occasions.

The decision tree for the 5 clusters also show that the affluence grade, having the mobile applications, participation in previous campaign, Facebook likes are important factors in targeting a certain group of customers.



**Appendix: Full Diagram**

