MKT-6223: Database Marketing
Group Assignment II

Instructor: N. Kumar, Ph.D.

All Data files are on the class web page.

1. **Mass Transit:** Through a survey of approx. 600 consumers, the Mass Transit Authority has collected information on consumers' preference for use of Public Transportation as a function of gasoline prices. In addition the Mass Transit Authority collected attitudinal/demographic data on these same consumers.

   There are two data files: (a) *calibration.xls* that has customer id their preference for using public transportation (*x1-x12*) along with their demographic characteristics (b) *prospect.xls* has customer id and demographic information pertaining to a list of prospective users. See data description file for additional details.

   The department is seeking your advice on whom to target. Base your decision on the following analysis:

   (a) Use the primary data set on usage intention (*calibration.xls*) to perform cluster analysis. Specifically, use variables *x1-x12* to perform this analysis and segment customers into two groups - potential *users* and *non-users* of public transportation.

   (b) Using the demographic variables in *calibration.xls* perform Discriminant analysis to identify the variables that best discriminate potential users from non-users. Use the Discriminant Function from the above to classify consumers in the *prospect.xls* into the appropriate groups (testdata option can help you do that in one shot).

   (c) Based on the results in (b) the Mass Transit Authority engaged in a campaign to promote the use of their mode of transportation amongst the 300 consumers in the prospect list. They tracked the behavior of these 300 consumers and created a variable *used* = 2, if the consumer used mass transit and is set to 1 otherwise. This information is available in *validation.xls*. Compare your predictions from (b) with the actual behavior of the prospects to see how well you did in classifying *users* and *non-users*.

Interpretation of the Results

(i) Based on your results in (a) how would you determine which cluster represents consumers that are *users* (*non-users*) of public transportation?

(ii) What criteria would you use to assess the goodness of your segmentation in (a)?

(iii) Based on your results in (b) which demographic variables help discriminate potential *users* of public transportation from *non-users*. Why?

(iv) How good is your discriminant analysis classification – what criteria would you use to ascertain this?

(v) Comment on the cross-tabulation in (c). What does that tell you about the effectiveness of your classification procedure.