**Module 2**

**Assignment 8**

This assignment will be addressing the concepts of one of the data reduction techniques called Principal Component Analysis (PCA). In the given dataset, the following variables are given. total population (Pop), median years of schooling (School), total employment (Employ), employment in health services (Health), and median home value (Home).

**Use the dataset provided “city.csv” and answer/execute the following questions/ tasks:**

**Questions/tasks:**

1. Perform principal components analysis to understand the underlying data structure
2. Create a scree plot to determine how much variation each principal component captures from the data
3. It is important to check for outliers because they can significantly influence the results. Use ‘Mahalanobis distances’ in your data and plot them to observe the outliers
4. Perform normalization on the given dataset and perform PCA and compare with the results which you got from PCA results which done using raw data.
5. Check the proportion of variances of the last two components in the result table. Are they significant when compared to other principle components?

**Duration**: 3 Hours