**BUS/CSC 328 Data Analytics**

**Homework #8 Clustering and Ensemble Modeling**

Step 0: Create a Word document that will act as the scientific record of your work.

Step 1: Import the Wholesale dataset and perform a complete inspection.

Step 2: Check for missing values and declare your strategy (imputation or removal).

Step 3: Create and run a recipe to prepare the data for support vector machine modeling.

Step 4: Run three different regression-based SVM models (linear, polynomial, and radial) on the full dataset and identify the model with the best performance. The dependent (response) variable is to predict how much Fresh food will be purchased based on the other predictors. Use a tune length of at least 10 for each model and let the algorithm choose the best hyperparameters.

Step 5: Without removing any variables, complete a model-based clustering analysis and show which model and how many clusters resulted in the best BIC value. Paste the cluster plot and printout into the Word doc.

Step 6: Add the cluster classification to the Wholesale dataset as a new variable.

Step 7: Filter the Wholesale dataset by cluster and run three different support vector machine models (linear, polynomial, and radial) for regression on each cluster. The dependent (response) variable is to predict how much Fresh food will be purchased based on the other predictors. Use a tune length of at least 10 for each model and let the algorithm choose the best hyperparameters.

Step 8: Select the best model for each cluster and show the metrics associated with those models.

Step 9: Did splitting the data into clusters allow higher accuracy than using the full dataset? Describe in detail what you found.