**Fordham University – Gabelli School of Business**

**Data-Driven Marketing Decisions**

**Prof. Mohammad G. Nejad**

**Project 2: Customer Segmentation Using Large Scale Transactional Data**

A specialty multi-channel catalog company has reached out to you identify different segments of their customers by analyzing a large data set. The company has an active presence on the Web and also mails a high volume of catalogs annually as its primary marketing vehicle. The company accepts orders via their website, phone, and by mail.

The company has provided you with a rich data set with over 226,000 records, reflecting over 137,000 orders from 100,000 customers. The data includes all orders, dollars, items, order lines, and returns for those customers 12/16/04 thru 9/17/12 over transactions in all channels (phone, web, and mail). These orders are from 100,000 random U.S. customers (representative of all their customers) with a first purchase date of on or after 12/16/2004.

Your task is to:

1. Pick a company and get the approval from the professor
   1. You can choose any company that sells via more than one retail channel such as online, phone, in-store, department stores, etc.
   2. Previous students have used companies such as J. Crew, Macy’s, Foot Locker, Lululemon, CVS, Puma, etc. You may not use these companies.
2. Assume that the dataset is from the retailer
3. Transform the transactional data to customer data and then pinpoint different customer segments of the retailer and discuss managerial implications.

Other tasks:

1. Design the data analysis plan
2. Conduct detailed analysis
3. Provide actionable strategies to executives and managers.
   1. Your strategies must reflect the reading assignments and in-class discussions we had on dealing with different customer segments
4. Communicate your analysis and results in the form of a professional report.
   1. Your report will include information required by people with different roles in a company including executives, managers, and data analysts.

**Questions to Ponder (no need to submit anything):**

1. Identify the primary and secondary goals that you are seeking to accomplish
   1. What is the unique contribution(s) of your project and what can a manager do with them?
2. The dataset contains transactional data which is not suitable for customer segmentation. The first step is to aggregate and transform the data to customer data which includes calculating and creating variables that will be useful in this analysis. What variables are you able to create by aggregating the data? Be a Creative Analyst!
   1. Variables to be used in cluster analysis and the ones for post-hoc analysis?
   2. Study the variable list, how can you calculate the variables and prepare the data?
   3. 3 to 5 variables for cluster analysis; no limit on variables for post-hoc analysis

**Analysis**

1. Watch the video titled “Video 4 - Hybrid CA - Putting everything together.”
   1. Follow the steps discussed in this video in all analysis discussed below.
2. Aggregate the data file and prepare the variables.
3. Conduct hierarchical cluster analysis
   1. Decide on the size of the subset of the main data file.
      1. What percentage of the entire data do you want to use?
   2. Decide on
      1. Standardizing variables;
      2. Dealing with outliers;
      3. Potential multicollinearity issues;
      4. Clustering method and distance measure.
      5. Is missing variable(s) a problem?
   3. Start the analysis and conduct HCA
      1. Try at least two random subsets and two HCA methods (Wards, average linkage, furthest neighbor) for each random subset.
      2. Try different numbers of clusters, etc.
   4. Identify your final seeding points and store them in a data file.
4. Conduct K-Means cluster analysis for each HCA result identified in step 3.d.
   1. Use initial seeds identified in step 3.d to run K-Means CA.
   2. Compare K-Means outcomes.
   3. Pick the final clustering solution.

**What to turn in?**

A report communicating your analysis, findings, and recommendations as described in the file “Project 2 Report Guidelines”. No need to turn in the answers to the questions/topics explained in this handout.