**Catalog Data – Mini Project 1**

The file Catalog Dataset.zip contains a dataset in plain text format and two additional files – one that describes the data layout for each record in the data file, and one that describes the variables (data dictionary). Note that some fields are not included in the dataset (as noted in the data dictionary). The last two fields mentioned in the layout file (CR, LF) are the “carriage return” and “line feed” characters used to mark the end of the record. SAS will automatically handle this correctly. You do not need to read these characters separately.

The dataset is from a catalog company that mailed catalogs seasonally to existing customers, customers of subsidiary/affiliated companies, and customers reached via web advertising. The catalog was promoted through both direct mail and email. The data Set contains twelve years of data, through April 30, 2009. Information includes catalog orders, order source, quantity of items purchased, returns, payment information, and the zip Code of the purchaser. There is one record per order, with multiple orders per household. Orders with the same household are indicated with matching Household-ID numbers (one number per unique household.) The file contains 14,448 order records from 10,000 unique households. The accompanying data dictionary explains the data in more detail.

In addition, there is US zipcode data that provides the city, county and state details for each zipcode in the US. This can be useful to aggregate the data by geographic territories. There is also demographics data by zipcode from the US 2000 census data in the zip file US 2000 Census Data.zip. This can be used to understand how demographics affects catalog orders.

Initial Preparation

* Import the dataset into SAS (see the data organization file).
* Make sure that numbers are stored as numbers and dates are processed correctly. Perform additional processing if necessary to ensure that the fields are in the correct format.
* Examine the data in each field of the dataset. Look for missing data. The objective is to understand the organization and meaning of the data better.

Data Inspection and Cleaning

* The data dictionary describes how orders, returns, cancellations, refunds, shipping etc. are handled. Is the data consistent with this description? If not, what are the discrepancies? How will you handle the discrepancies (if any)?
* The data dictionary mentions that refunds, cancellations, returns and other order information are handled differently depending on the transaction date. Update the dataset such that the data has the same consistent interpretation regardless of the date.
* Are there other variables in the dataset that require inspection and cleaning (to ensure consistent interpretation or avoid missing data)? Are there other variables that are interdependent and need to be checked for consistency? Are there data fields that you can / should drop?

Exploratory Analysis

* What are the different ways in which you can categorize orders? What variables are useful and reliable for this purpose?
* What are different ways in which you can categorize customers?
* Are there any interesting patterns or trends that you see in the data? Over time? Across geographies?

Developing Marketing Insights

* Brainstorm and generate a list of marketing ideas / questions / insights relevant to this dataset. These marketing ideas / questions/ insights should help a marketing manager take some marketing actions to improve the business. So for item on your list, you need to be clear how or why it would be useful for a marketing manager.

REMEMBER: Not every analysis or hypothesis test you can perform with the data is useful to know for a manager. So think carefully about why a certain question should be interesting or important.

* Frame each item on the list as a question that you can answer using the data.
* For three of these, conduct the necessary analysis to answer the question using the data. Consider whether there is more than one way to pose the question and/ or conduct the analysis.

Presenting Insights

* Prepare a 10-min presentation to discuss the three insights that you obtained from the data.

REMEMBER: Not every statistically significant result, is actually important for a manager. Think carefully about whether or how that piece of insight will affect marketing decisions. For e.g., even if household income has a significant in fact on sales, if the actual magnitude of the effect is not that large (given the range of possible household income), then this finding is not of much consequence.

* In your presentation, describe each item, explain why it is important or interesting for marketing, then present your analysis and conclusions, and finish by making recommendations for specific marketing actions that the firm can take.

Deliverables

* Project Report
  + What changes / assumptions you made while cleaning the data
  + List of marketing ideas you identified. Why these are interesting or important for amarketing manager. Exploratory analysis, if any, that led to these ideas.
  + The analysis, results and conclusions for three of the ideas that you chose for your presentation
  + Include all SAS code in the Appendix.
* Project Presentation as discussed above.