# ANA 610 Homework #1

**Variable Annuity Dataset**

The marketing department of a large financial institution wishes to conduct a targeted marketing

campaign for the bank’s money market product. The VP of Marketing has sent a request to your analytics team to build a scoring (predictive) model to be used to score current bank customers (who do not have a money market account) for their likeliness to be receptive to such an offer.

The marketing department has identified the following qualifications to be used for the target and non- target samples of customers (**assume analysis is taking place in June 2015**):

* at least 1 product out of checking, mortgage, credit card, insurance, savings or IRA
* 18 years old as of June 1, 2015
* at least 1 year of tenure as of June 1, 2014
* FICO score >= 650
* either has or does not have money market account

Based on input from the marketing department, the IT department has made available the following 19 files for your analysis:

* (SAS) Base client file: client\_hw
* (txt) Credit Bureau file: credit\_bureau\_hw
* (SAS) Insurance Product: client\_ins\_act\_hw, ins\_act\_hw
* (SAS) Checking Product: client\_chk\_act\_hw, chk\_act\_hw
* (SAS) Mortgage Product: client\_mtg\_act\_hw, mtg\_act\_hw
* (SAS) IRA Product: client\_ira\_act\_hw, ira\_act\_hw
* (SAS) Credit Card Product: client\_cc\_act\_hw, cc\_act\_hw
* (SAS) Savings Product: client\_sav\_act\_hw, sav\_act\_hw
* (SAS) Money Market Product: client\_mmk\_act\_hw, mmk\_sav\_hw
* (SAS) Client Address: client\_address\_hw
* (SAS) Census data: census\_2010\_z5summary
* (csv) State FIPS data: state\_fips Your tasks are the following:

**Task #1 (60 pts):** Generate a data audit report (using the audit report template) to be shared with both the marketing and IT department; include a check of the available modeling sample size after application of target qualifications.

**Task #2 (20 pts):** Assemble all operational tables into 1 modeling table.

Note:

* The data are for the last 3 months prior to June 1, 2015
* From the Census file append State, MedianHValue, MedianHHInc, PctOwnerOcc and the ZIP Code variable (ZCTA5). Perform the data audit only on this reduced file.

# Donor Solicitation Dataset

The file “Data Dictionary – Donor.pdf” presents a 19,000-record dataset of individuals who had been contacted in a veteran’s donation campaign in 1997. Some gave; some did not. We will explore this dataset over the course of this class. The ultimate objective to prepare the dataset so it can be used to build a predictive model of donor likelihood which will be used to score a much larger mailing dataset for a new solicitation campaign.

**Task #3 (40 pts):** Perform a data audit on this dataset: S\_PML\_DONOR\_HW

**Task #4 (20 pts):** Perform some basic integrity checks on the dataset: S\_PML\_DONOR\_HW

1. Check for duplicate records (show your code);
2. Refer to your data audit…does anything catch your eye with regards to the numeric variables?

Explain.

1. Refer to your data audit…same question with respect to the categorical variables.

# Task #5 (60 pts):

In examining the Donor dataset, you realize that there are no dates. Yet you know that you will be asked to generate reports based on the following dates: date when individual was entered into the file, date of an individual’s first gift; and date of an individual’s last gift.

1. Create these fields on the dataset, giving them a format of MM/DD/YYYY (HINT – assume you are conducting this analysis on June 1, 1997 and each month has, on average, 30.4 days))
   1. What is the date of the first file entry?
   2. What is the date of the first gift?
   3. What is the date of the last gift?
2. Create 3 additional fields, showing the YEAR (in YYYY format) for first file entry, first gift and last gift.
   1. In which year were the most individuals added to the file?
   2. Which year showed the largest average LAST\_GIFT\_AMT?
   3. What has been the trend in the number of people donating, in terms of LAST\_GIFT\_AMT?
   4. In which STATE and YEAR was the average LAST\_GIFT\_AMT $57.50?

# Homework deliverables:

* Task #1, #2 and #5:
  + 2 data audit reports (Word doc)
  + 2 SAS modeling tables
* Task #4 and #5
  + Word doc with your analysis and discussion, including all tables and charts