## Assignment 9

This is an individual assignment focused on using logistic regression in Python.

The data is loosely based on this Kaggle contest and can be found here:
https://s3.amazonaws.com/programmingforanalytics/repeater_prediction.csv

We are helping a chain of grocery stores to predict if sending out a personalized coupon to a customer will make him buy the product on the coupon. The target variable (dependent variable) is *repeater* (1 means the coupon was a success, 0 means the coupon was a failure). All the other variables are independent variables and can be used as features in the model.

1) Import the data into a Pandas data frame and perform data exploration, such as summary statistics of variables and frequency tables. Include some plots where appropriate
2) Decide which variables should be continuous variables and which should be factor variables. Check the types Pandas automatically assigned by using pandas.DataFrame.dtypes. Change the type of some variables, if you feel this is necessary, by using pandas.DataFrame.astype

3) Split the data into 70% train and 30% test set. Train a logistic regression with all the independent variables on the training set

4) Perform 10-fold cross validation, using *accuracy* as your scoring method