Homework 1, Business Analytics

Chengjun Liu, 7116099552

* Understand situation and need: we try to find more high-value customers by guessing what such customers are like from previous samples and judge the other potential customers.
* Data
  + The data are mainly collected from internal resources. Customers’ accounts information and purchase history will give different variables, revenues and costs. If possible, customer online activities, comments and surveys will provide more data on variables.
* General approach: We try to find variables that influence customer value. We use statistical models to judge how they influence values and use variables to predict potential customer values.
* Procedures
  + Understand the characteristics of my best customers
    - We will define the values of customers by building a formula. We use Revenue (User Payment) minus Cost (Mainly average product cost and service cost) to measure the customer value. We will present values of customers as a single number and the largest numbers correspond to our best customers.
    - We select and focus on several main characteristics of customers. Here we suggest using Single/Family, Value Growth, Gender, Billing, Age, Channel, Region, Race, Browsing Time. If possible, we can add more variables.
  + Find influential variables
    - Let variables in previous time period (like 6 months) be the influential variables and “current” customer values be the influenced result we want to learn. By comparing whether different factors make different value distributions in each variable, we can select the variables that indeed have influence. We can select those variables as the characteristics of our best customers.
  + Find more high-value customers
    - We match the characteristics of best customers with other current customers. It’s better for us to build a model to convert characteristics into predicted values. If previous characteristics make high values, we assume they are potential high-value customers.
    - We gather variables and customer values of different customers and try to build a proper statistical model. We can slice the time period of each customer into multiple samples to enlarge sample base.
    - For statistical model, we suggest neuron-network statistical model. After training the statistical model with high-value customers we observed and testing its reliability, we can input the characteristics of current customers and sort predicted customer values to find potential best customers.
    - We can observe the leverage of variables in the statistical model. If new customers and registered visitors match the characteristics that have a bigger leverage and score higher predicted values, we can target them as high-value customers.