SCRIPT

Team: MAASK

Objective:

Our main objective is to identify factors that convert a lead to a sale.

Approach:

We started by combining the "Leads" and "transactions" tables. We created a dealer-id that was based on the assumption that only one particular dealer was located for a given make and zip-code. Subsequently, the leads data-set was filtered out to contain only those cases for which dealers were present in transaction data (we assume here that if a dealer submitted one transaction, he would be diligent to submit all his related transaction).

We did some basic data manipulation and cleaning for all the tables, filtering the outliers. One of the variables MSRP was missing for 1 million records out of the total 1.2 million records and we imputed the missing values through manipulation and regression. We divided our final data set into test and validation frames and ran logistic regression on it. Our dependent variable is whether the sale took place or not after a lead was submitted.

Results

What qualifies as success is when there is a lead submitted for a given model and the customer ends up buying a car of the same model and make or one of a higher version.

SIGNIFICANT INDEPENDENT VARIABLES:

1. Dealer_conversion_type (very high, high, medium, low, very low)
2. Other vehicles considered – numerical
3. Level of interest
4. No. config. tried – for a particular make, model
5. Total config. tried – including everything
6. Promise_flag
7. New_used_flag

C-stat for the model was 0.83

Next steps:

a) Map a customer to the right type of dealer

b) Compare model performance using SVM, decision trees, neural-networks

c) Further analysis on the identified significant variables above.