Advertisement Selection and Predictive Analysis

Tracy Song-Brink

Jul 25, 2018

Customer and Conversion Data

- Conversion_data contains information regarding the customers, advertisements they have seen, and a flag identifying if they've converted.

- Customer_data contains demographic information about the customers.

Overview

- Background: Out of 499965 customers, the client randomly selected 145790 customers to test out 3 advertisements and recorded whether the customers were converted.

- Process:
  - Hypothesis testing on Advertisement conversion rates
  - Advertisement based predictive modeling

- Solution: Advertisement choice recommendation on the entire customer population

- Tools: Base SAS and SAS Enterprise Miner

Customer Data

- Total number: 499965

- Average age: 35.3 years old

- Region:
  - 67% in the US (11 States)
  - 18% in Canada
  - 11% in England
  - 4% in Japan

- Average Income: 107462 (dollars?)

- Gender ratio: 51% male and 49% female

U.S. Customers

count | 21347 - 21501 | 25601 - 25687 | 25697 - 25852 | 30018 - 81172
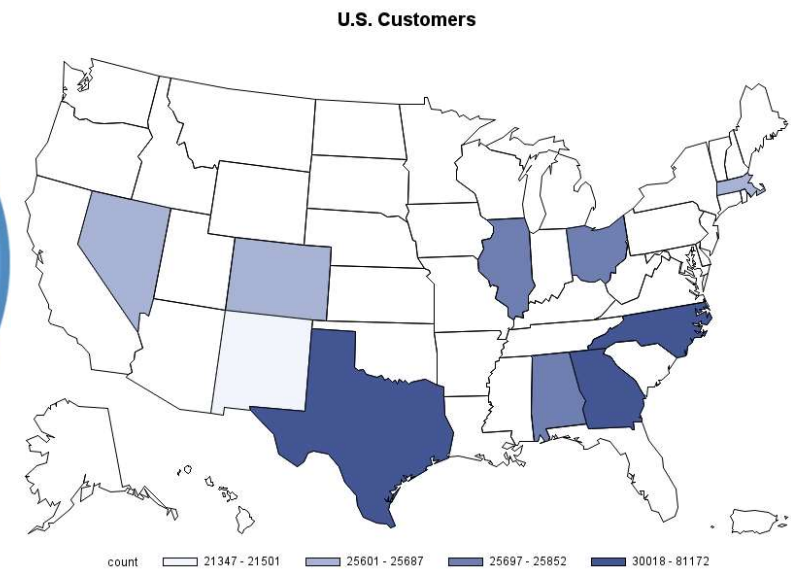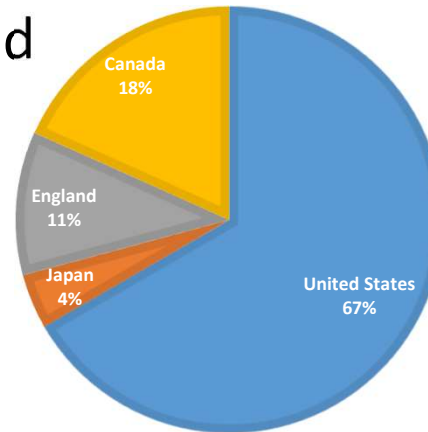
Conversion Data

- Details of the data:
  - Number of customer: 145790 (29% of 499965)

  - Overall Conversion Rate: 1.64%
    - Advertisement_id 26*: 1.90%
    - Advertisement_id 61*: 1.52%
    - Advertisement_id 75*: 1.50%
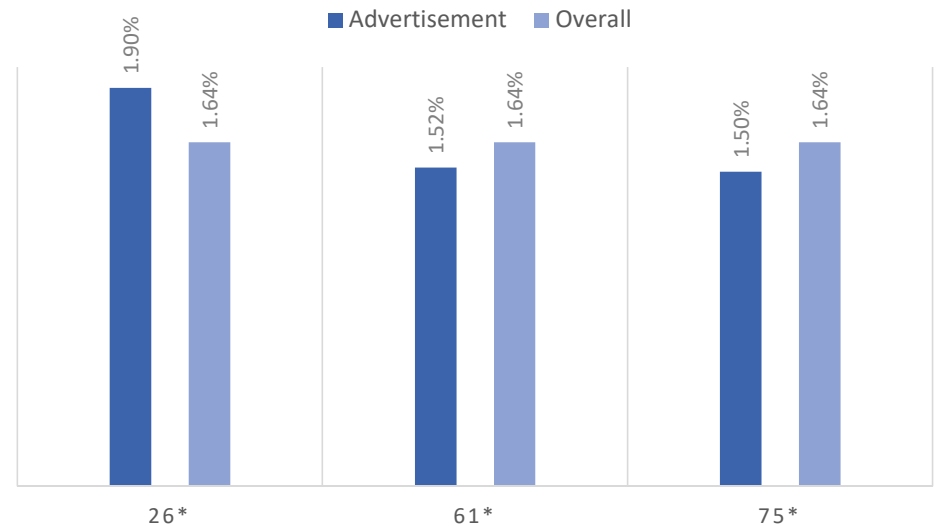
Is Advertisement 26* better the other two?

- Test of Hypothesis with Binomial Distribution
  - H_0: All three advertisements have the same conversion rates
  - H_a: conversion rates are different by advertisement
  - Confidence level 95% ($\alpha=0.05$)
  - The small p-value indicates rejection of the null hypothesis in favor of the alternative that Advertisement 26* is better than the other two.

Conversion Rate by Advertisement

The FREQ Procedure

conversion	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	47591	98.10	47591	98.10
1	923	1.90	48514	100.00

Binomial Proportion	
conversion = 1	
Proportion	0.0190
ASE	0.0006
95% Lower Conf Limit	0.0178
95% Upper Conf Limit	0.0202
Exact Conf Limits	
95% Lower Conf Limit	0.0178
95% Upper Conf Limit	0.0203

Test of H0: Proportion = 0.0164	
ASE under H0	0.0006
Z	4.5531
One-sided Pr > Z	<.0001
Two-sided Pr > |Z|	<.0001

Sample Size = 48514

6

Modeling

- Advertisement-based Modeling
- Models: Logistic Regression, Decision Tree, Neural Network
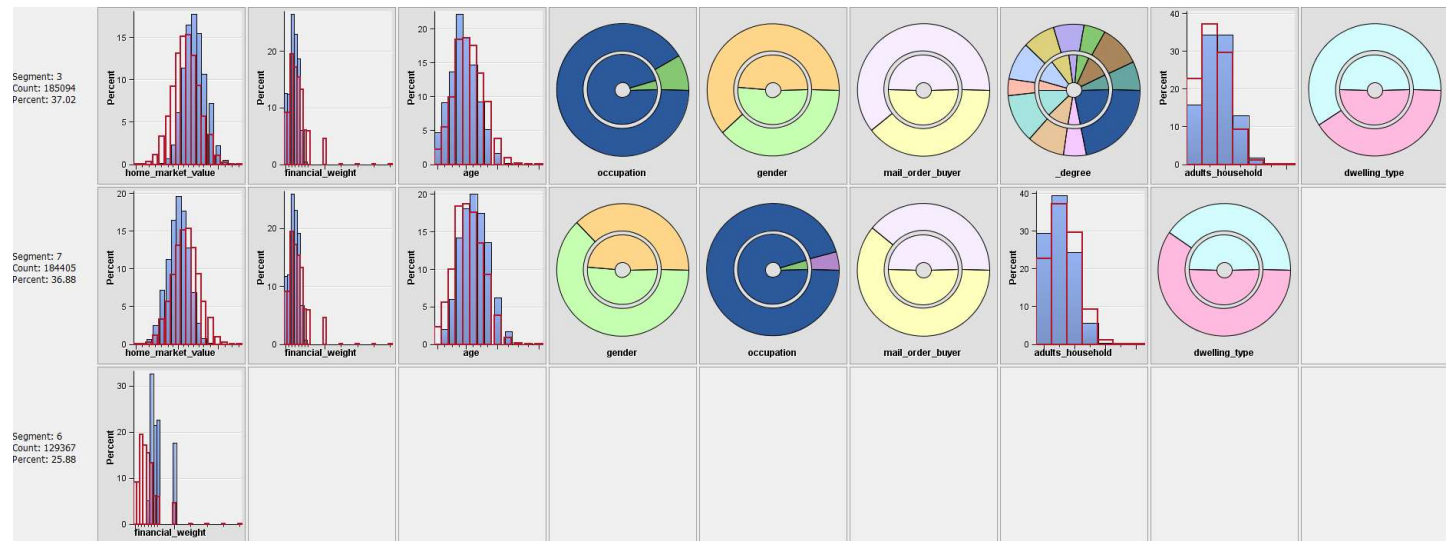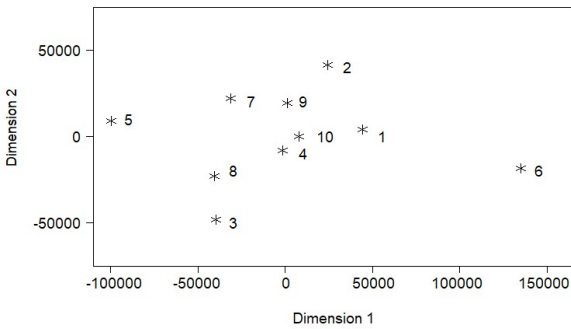- Data: Oversampling 60/40
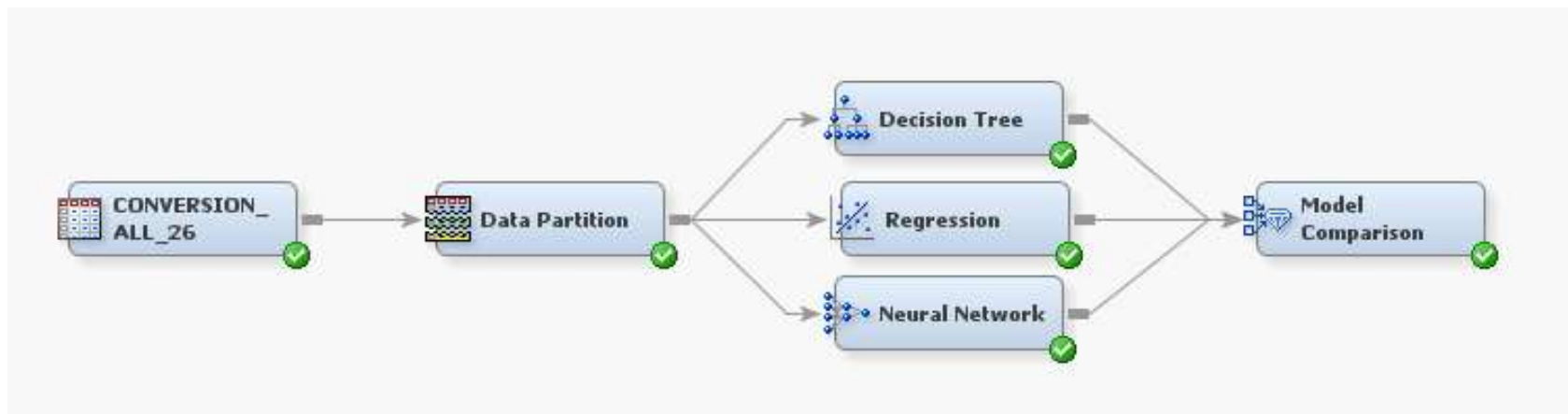
Scoring and Advertisement Choice

- Predicted Conversion Rates:
  - Advertisement 1: 4.32%
  - Advertisement 2: 4.90%
  - Advertisement 3: 4.12%
- Advertisement Choice Recommendation:
  - Advertisement 26*: 20954
  - Advertisement 61*: 23641
  - Advertisement 75*: 18758
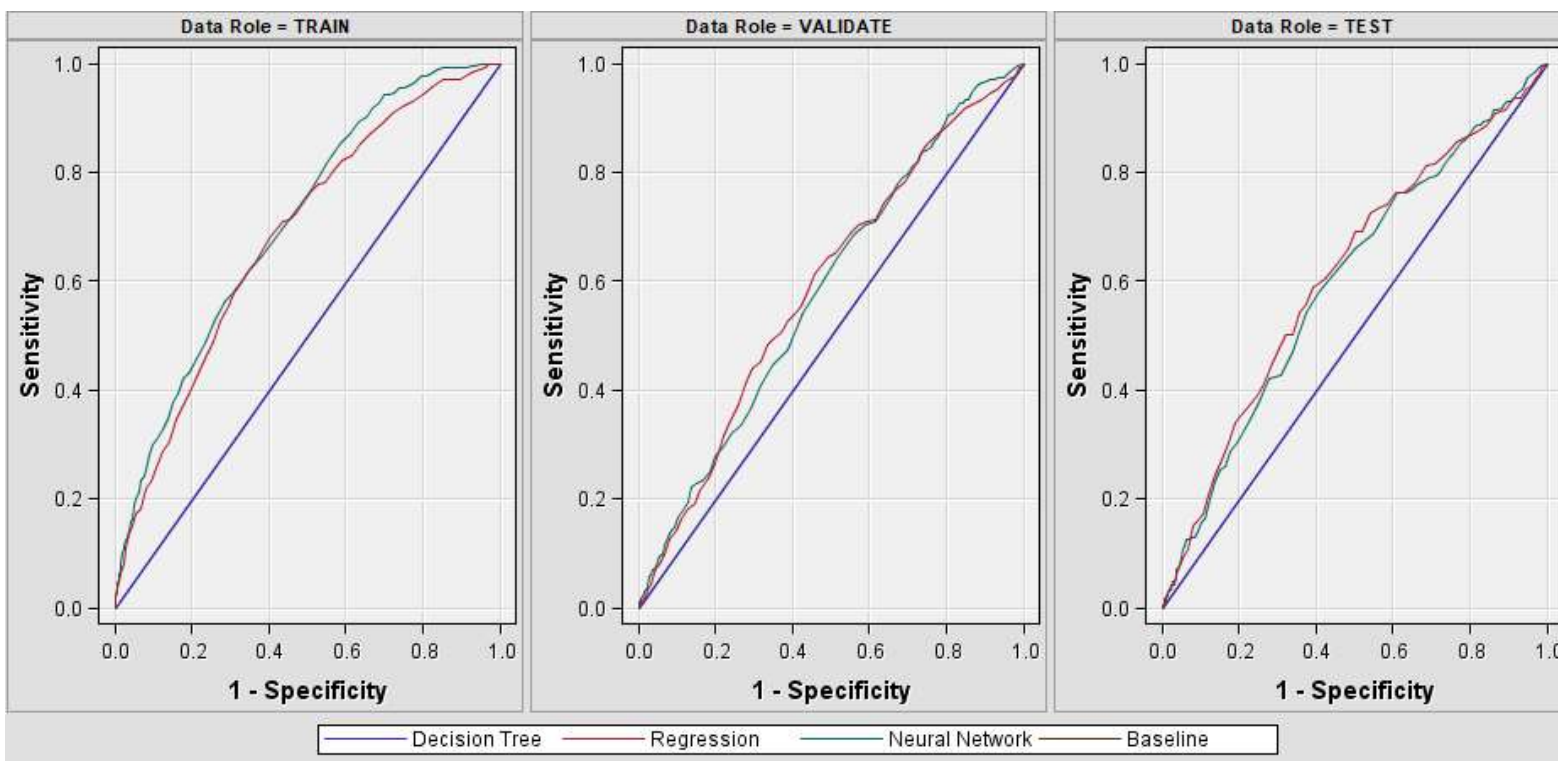 - No advertisements or random choice: 450401

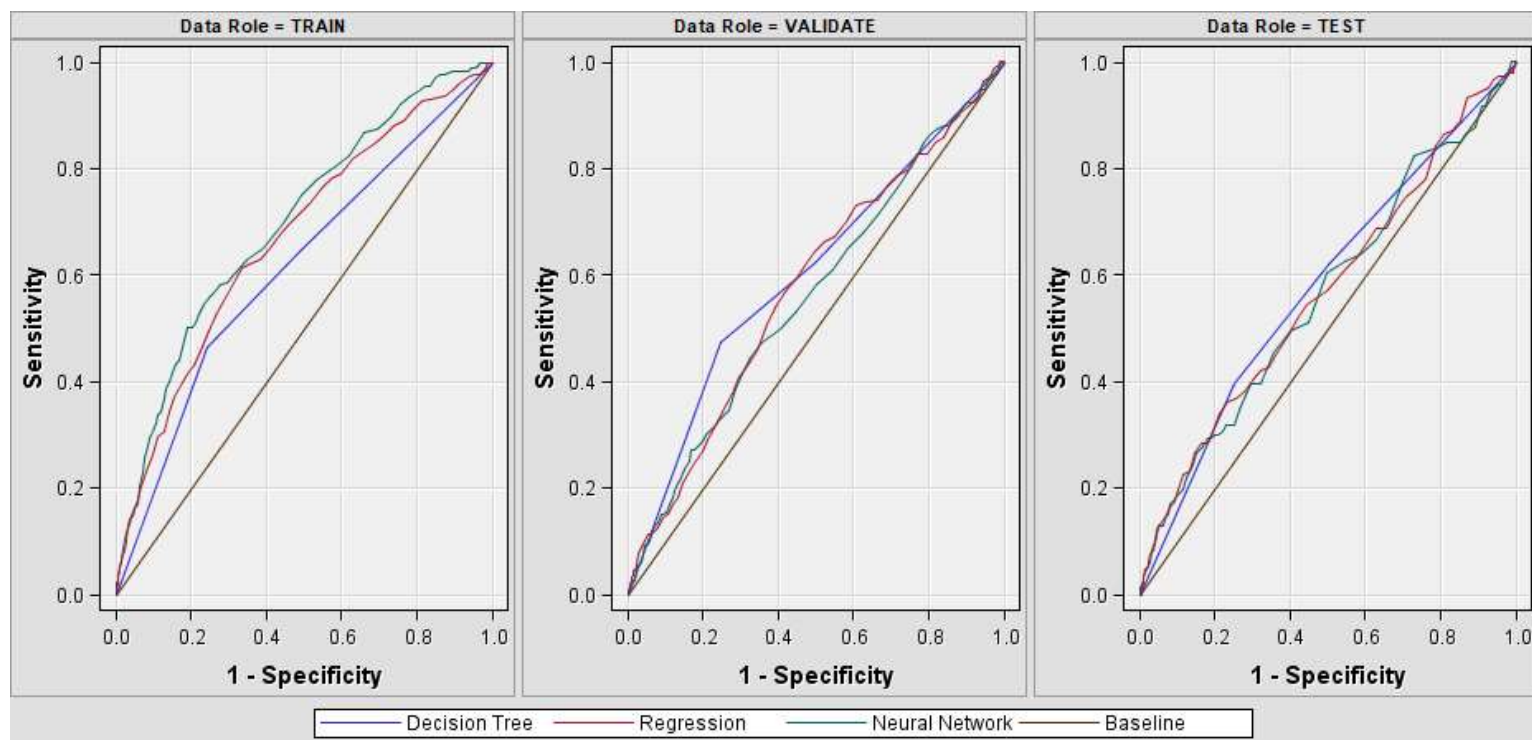# Appendix 1 : Customer Segmentation
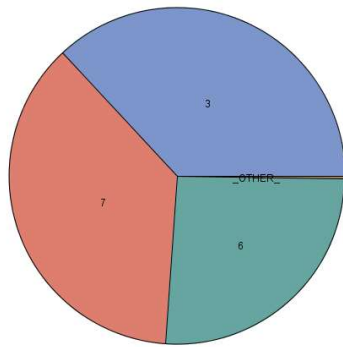
Appendix 2 : ROC of Advertisement 26*

Appendix 3 : ROC of Advertisement 61*

Appendix 4 : ROC of Advertisement 75*