Northeast
Member Segmentation Analysis

Created By: Byron (Pi-Jen) Tang

Agenda

- Background
- Objectives
- Approach
  - Data Preprocessing
  - Data Understanding
  - Problem Definition
- Methodology & Analysis Result
  - Revenue
  - Cost
  - Segmentation
- Recommendations
- Appendix

AAA Northeast

EMERITUS

COLUMBIA | ENGINEERING
EXECUTIVE EDUCATION

Background

AAA Northeast is one of the regional clubs comprising the American Automobile Association, covering Rhode Island, Connecticut, Massachusetts and portions of New York and New Jersey. AAA Northeast offer services such as roadside assistance, maps, and various discounts as part of their services.

- Roadside assistance is a costly benefit, particularly towing. Members who frequently use roadside assistance are less desirable.

- AAA also offers other paid services at highly competitive prices. They also offer insurance, travel and banking/loan products. AAA would like to increase the penetration of these services.

AAA Northeast EMERITUS COLUMBIA | ENGINEERING EXECUTIVE EDUCATION

Objectives

Provide a market segmentation of AAA members at household level for AAA Northeast to better serve their members. This analysis would allow AAA to:

- Better anticipate the needs of members
- Customize communications and offering to various segments
- Expend more effort driving acquisition and renewal of desirable members

AAA Northeast EMERITUS COLUMBIA | ENGINEERING EXECUTIVE EDUCATION

Approach

- Data Preprocessing
- Data Understanding
- Problem Definition

Data Preprocessing

Data Cleaning:

- Remove cancelled members
- Transform Income/Credit Ranges/Number of Children to numeric data

Feature Engineering:

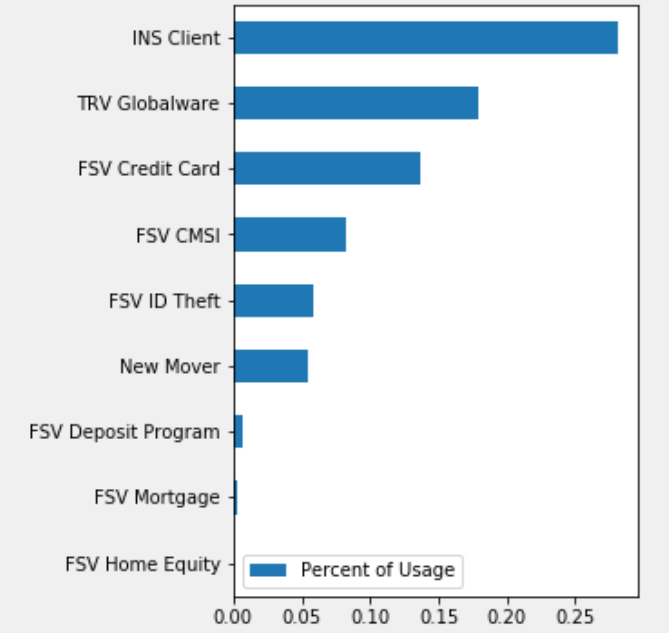
- Aggregate cost by year to create new cost variables – cost in 2014, cost in 2015, and so on
- Count number of member type by HH

Granularity – Household Level:

- Use sum or mean of numeric data
- Use mode for non-numeric data on variables that are consistent within HH
- Fill in missing value with median

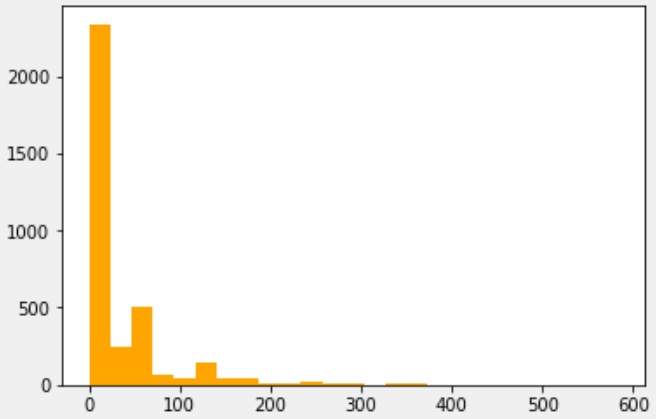
Data Understanding

Revenue



In general, AAA products have low market penetration. The highest penetration is less than 30% of the household.

Cost



The distribution of cost is highly skewed, with 67% households didn't generate any costs in 2019.

Segmentation

No. of Product Purchased	No. of Household
0	1589
1	1220
2	539
3	135
4	25
5	3

Around 45% of the households do not use any products from AAA Northeast, showing big market opportunities

AAA Northeast

EMERITUS

COLUMBIA | ENGINEERING EXECUTIVE EDUCATION

Problem Definition

Predict Probability of Purchasing a Product

Goal: Find the best models that give the most precise probability of purchasing

Challenge: Unbalanced data

Solution: Up-Sampling

Process & Model (For Each Product):

- Up-Sampling
- Cross Validation
- Decision Tree Model Optimization
- Bagging + Optimized Decision Tree

Forecast Cost in the Next 12 Months

Goal: Find the best models that give the most precise probability of generating cost

Challenge: With skewed distribution, cost prediction has low performance (RMSE)

Solution: Modify question and predict whether a customer will generate cost

Process & Model:

- Cross Validation
- Decision Tree Model Optimization
- Bagging + Optimized Decision Tree

Explore Market Opportunities

Goal: Explore clusters that would give actionable HH segmentations and enable product strategy

Process & Model:

- Gather data used for segmentation:
  1. Probability of purchasing each product
  2. Probability of generating cost in the next 12 months
- Apply K-Means Clustering

AAA Northeast EMERITUS COLUMBIA | ENGINEERING EXECUTIVE EDUCATION

Analysis Result

- Revenue
- Cost
- Segmentation

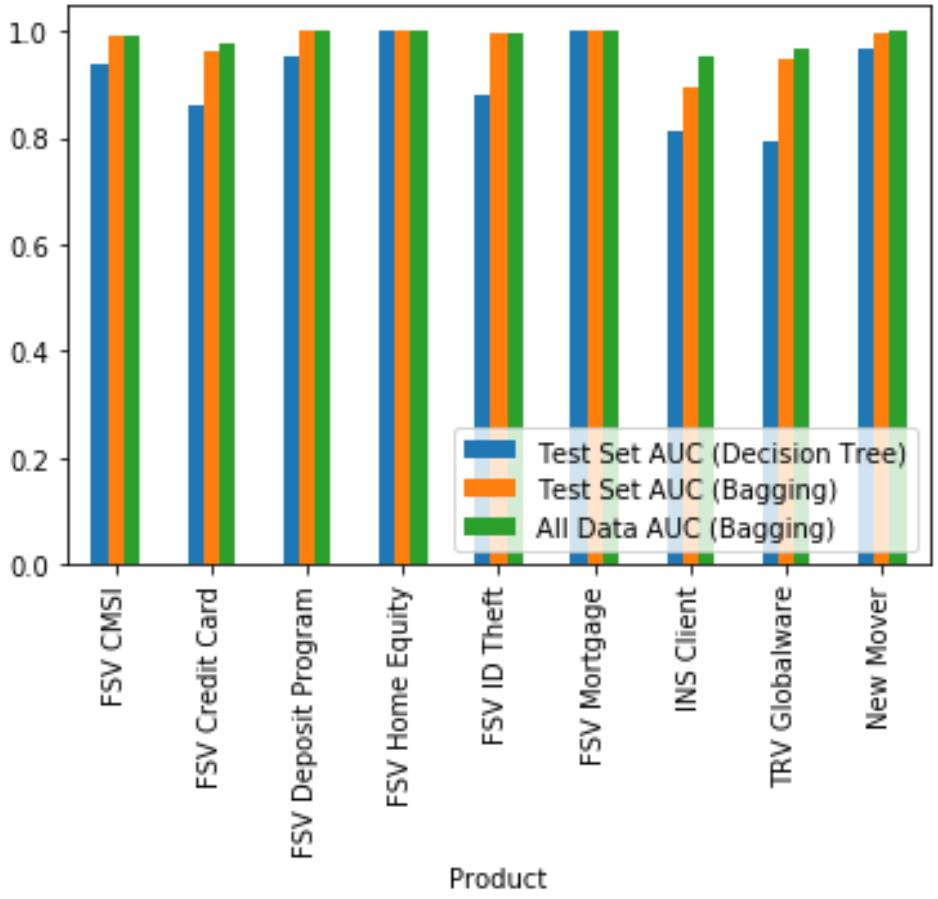
AAA Northeast

EMERITUS

COLUMBIA | ENGINEERING
EXECUTIVE EDUCATION

Revenue

Predict Probability of Purchasing a Product



Result – Average AUC:

- Test Set (Decision Tree) 0.911474
- Test Set (Bagging): 0.975726
- All Data (Bagging): 0.986834

With high AUC on predicting all products, we can be confident that the probability of purchasing would be a good reference on potential buyers for each product.

AAA Northeast EMERITUS COLUMBIA | ENGINEERING EXECUTIVE EDUCATION

Cost

Forecast Cost in the Next 12 Months

Time-Relevant Variables: Use the cost before year 2019 as year n-1, n-2, and so on to predict the cost in year 2019. Afterwards, treat year 2019 as year n-1, year 2018 as year n-2, and so on to predict the cost in year 2020.

Assumption: The household info would remain the same in year 2020

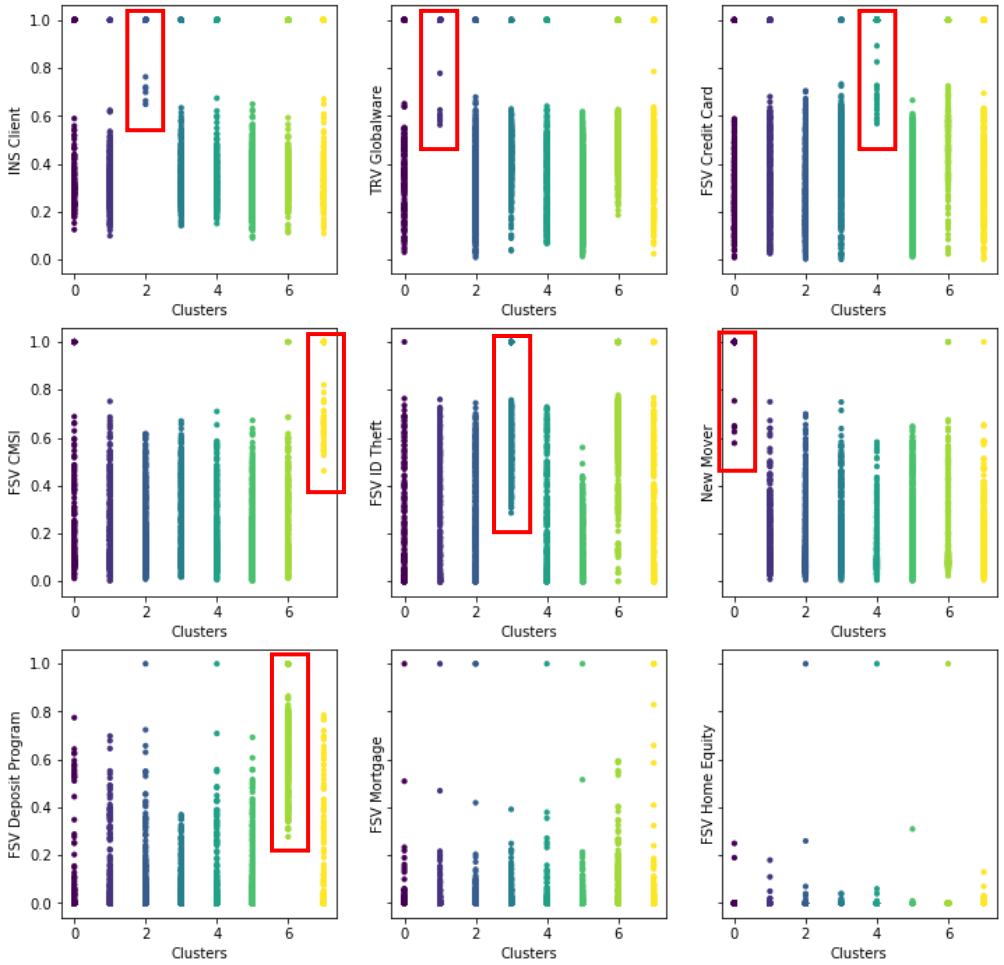
Result: The best model, bagging + decision tree, returns fair result to get probability of generating cost.

	Accuracy	AUC
Test Set	0.77	0.77
Whole Data Set	0.76	0.80

AAA Northeast EMERITUS COLUMBIA | ENGINEERING EXECUTIVE EDUCATION

Segmentation – Number of Segments

Explore Market Opportunities



When the households are clustered into 8 or 9 groups, we can identify target groups with high interest on a particular product for 7 products.

Therefore, it seems 8 is an ideal number of clusters.

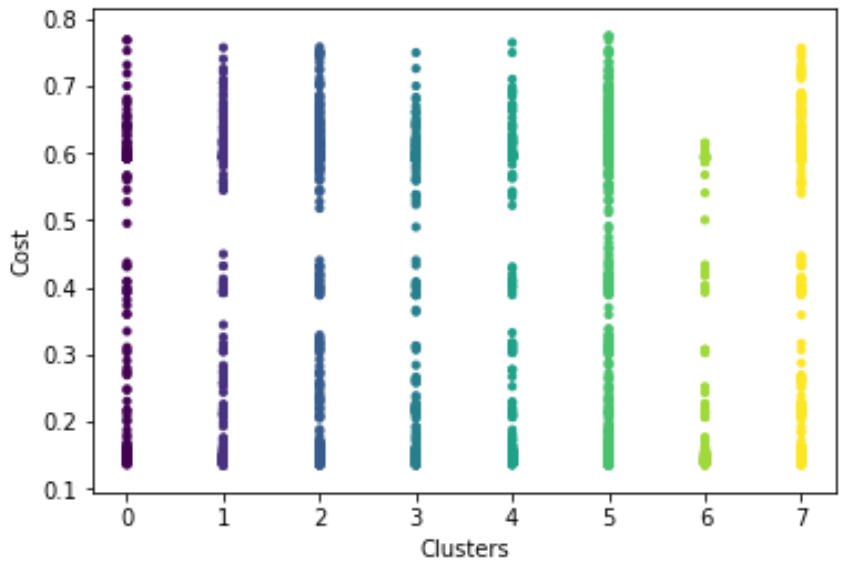
AAA Northeast

EMERITUS

COLUMBIA | ENGINEERING EXECUTIVE EDUCATION

Segmentation – Cost across Segments

Explore Market Opportunities



Predicted probability of cost does not have noticeable difference across clusters. One reason might be that cost is generated randomly as customers only need road service in emergency. It is also possibly because cost is hard to predict.

AAA Northeast EMERITUS COLUMBIA | ENGINEERING EXECUTIVE EDUCATION

Recommendations

1. Use the segmentation result to target HHs by product and make tailored product strategies. (See table)
2. Focus customized product strategy on top 7 popular products
3. Collect more information for cost prediction or conduct more exploration analysis on the pattern of cost.

- The table summarizes the size of potential HH for each product.
- Each segment represents a target group that is much more likely to purchase the product than the others.

Product	Targeted Household Size*	Avg Prob. of Purchase (Target HHs)	Avg Prob. of Purchase (Non Target HHs)
INS Client	6	0.701071	0.419056
FSV Credit Card	19	0.672262	0.352905
TRV Globalware	7	0.619258	0.391825
FSV CMSI	48	0.638845	0.185443
FSV ID Theft	281	0.544852	0.218673
New Mover	6	0.707800	0.172453
FSV Deposit Program	361	0.616487	0.055583

* The households that have purchase the product are excluded.

AAA Northeast EMERITUS COLUMBIA | ENGINEERING EXECUTIVE EDUCATION


THANK YOU

Appendix

Performance of other models tried for product prediction, using 'INS Client' as an example.

kNN

```
Best Number of Neighbors: {'n_neighbors': 9}
Accuracy on Training Set: 0.5888475836431226
Accuracy on Test Set: 0.5966303270564915
AUC: 0.6263004871915763
```

Logistic Regression

```
Best Parameters: {'max_iter': 1000, 'tol': 10, 'C': 0.0001}
Accuracy on Training Set: 0.5563816604708798
Accuracy on Test Set: 0.5441030723488602
AUC: 0.5618890460474619
```

Random Forest

```
Best Parameters: {'n_estimators': 150, 'max_features': 'auto', 'max_depth': 5}
Accuracy on Training Set: 0.6042131350681537
Accuracy on Test Set: 0.6164519326065411
AUC: 0.6562293729372938
```