AAA Northeast – Capstone Project

Post Graduate Diploma in Machine Learning and Artificial Intelligence

Columbia Engineering

By – Minesh Chunawala

Date – 19/12/2019

CONTENTS

1. Background
2. Objectives
3. Methodology
4. Exploratory Data Analysis
5. Predict Probability of Purchase
6. Predict Roadside usage and Costs
7. Market Segmentation
8. Key Recommendations
9. Appendix
10. References

1. Background

AAA Northeast is one of the regional clubs comprising the American Automobile Association. AAA Northeast covers Rhode Island, Connecticut, Massachusetts and portions of New York and New Jersey. The American Automobile Association has 58 million members in the US and Canada. They offer services such as roadside assistance, maps and various other financial services.

Roadside assistance is a costly benefit, particularly towing. AAA NE would like to understand drivers behind roadside service and costs. AAA also would like to segment their members that will enable them to extend other financial services.

Dataset - AAA NE has provided the dataset of customers predominantly from Rhode Island and Connecticut. Data consists emergency road side events. Each event has critical information about the household, member and type of incident.

Problem Statement:

1. AAA NE would like to understand what contributes to high ERS costs. Use the roadside data to make recommendations to increase revenue and reduce costs. Make predictions of roadside usage and costs for the next 12 months.

2. In spite of having large customer base for Roadside services, AAA NE doesn’t seem to have similar market penetration of their other services e.g. credit card, mortgage, insurance. Analyse current customer data and predict the probability of buying these products.

3. AAA NE would like to perform market segmentation of their members to understand consumer behaviours and accordingly make business decisions.

2. Objectives

Provide a market segmentation of AAA members (or member households) that allows AAA Northeast to better serve their members. They would use this analysis to:

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

3. Methodology

1. Data Cleansing and Exploratory Data Analysis.
2. Understand dimensions that drive revenue and costs.
3. Feature Engineering.
4. Predictive Model Building.
    1. Logistic Regression Model.
    2. Gradient Boosting Regression Model.
5. Use K-Means Clustering for Segmentation.
6. Analyse findings and make recommendations.

- **Logistic Regression Model** – Simple model to predict binary output. Ensemble method were tested too.
- **Gradient Boosting -** is a method of converting a set of weak learners into strong learners by optimising the loss function.
- **K-Means Clustering** – most popular clustering method. It partitions data into k clusters using mean distances.

4. Exploratory Data Analysis

1. Explore data features

2. Process member data for null values and outliers.

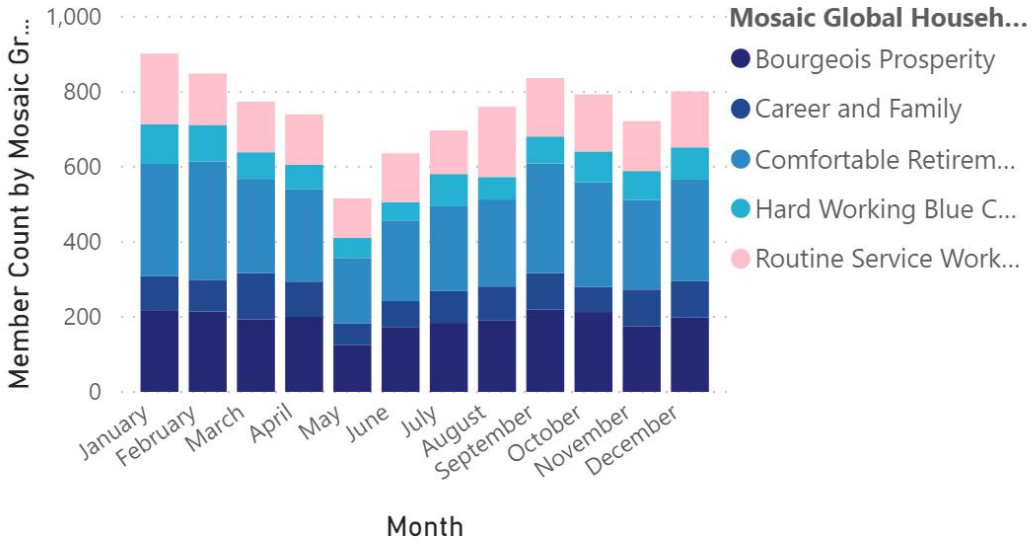
3. Select the most features relevant for model.

4. Summarise the data at a household level.

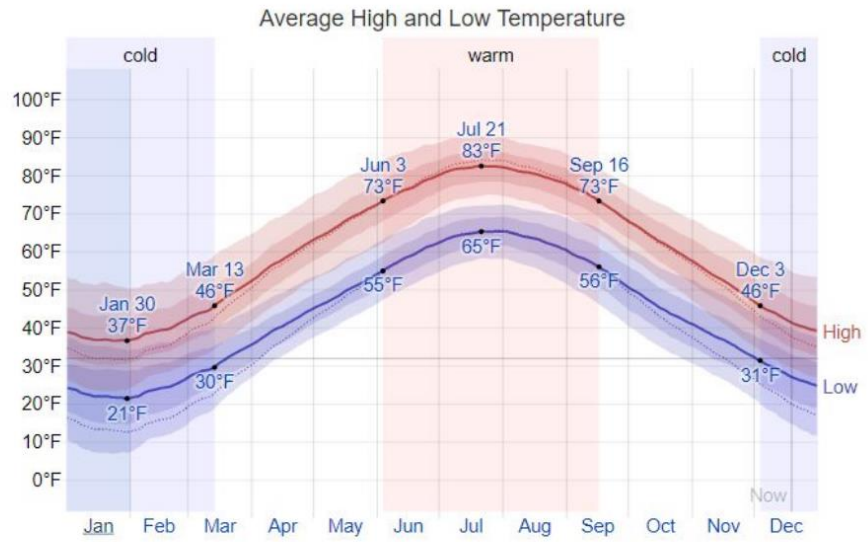
5. Feature Engineering
    1. Use findings from the household level data.
    2. Use correlation matrix to understand correlation between features.
    3. Use OneHot and Label Encoding to deal with binary and multi-categorical data.

Exploratory Data Analysis

ERS Calls by Month & Mosaic Group



Providence Weather - Monthly



Findings

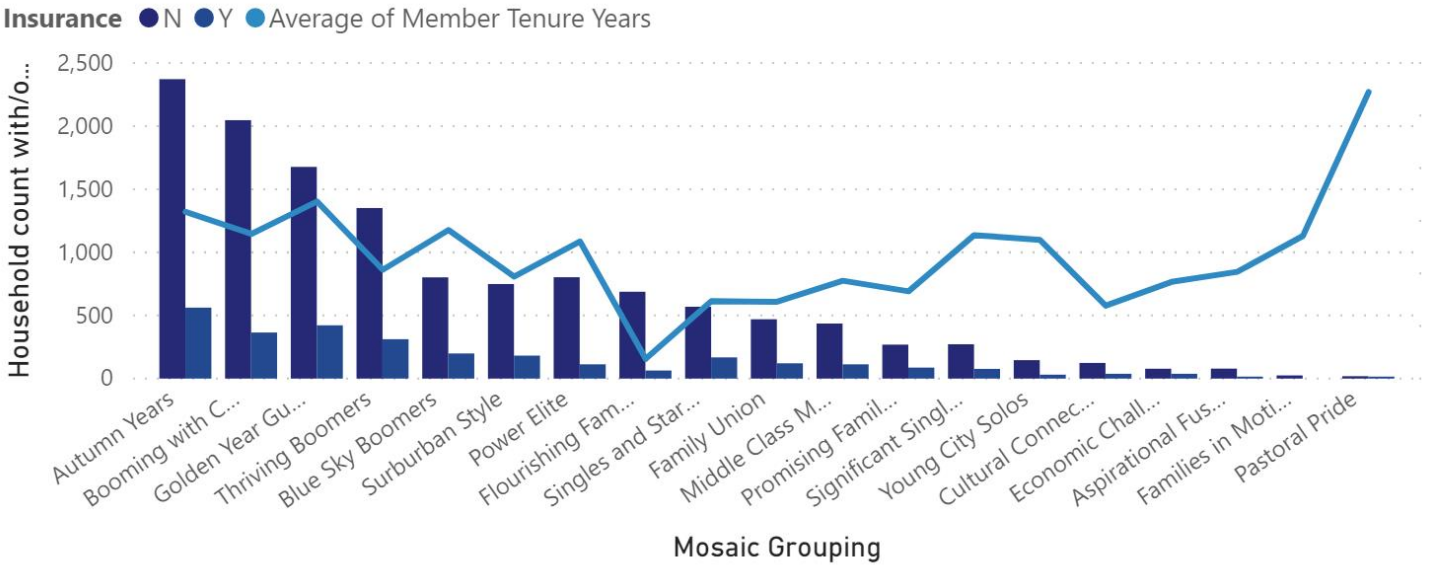
1.Elderly members seem to need ERS more.

2. Severe weather conditions is a major contributor.

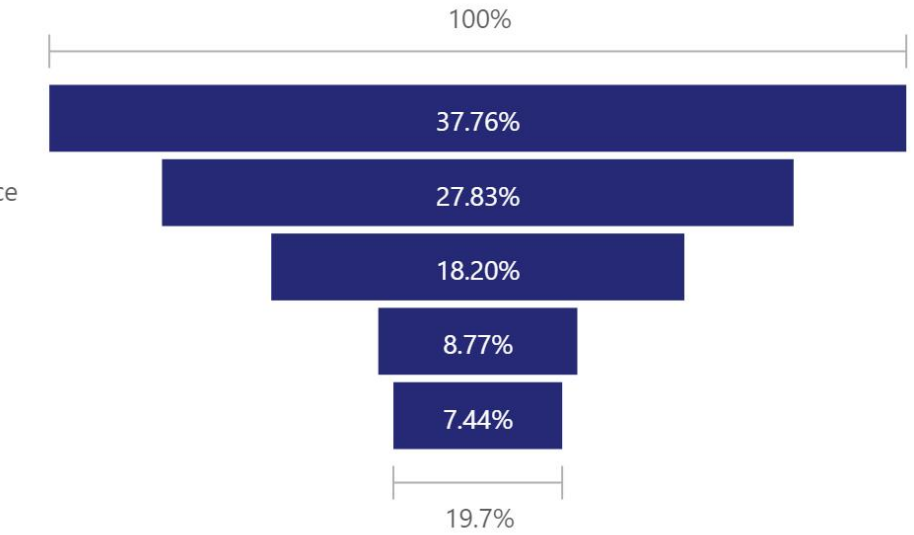
3. Aged and Middle Aged Family with long tenure are more likely to buy other financial services.

4. Towing, Battery Service and Flat Tire are major cost contributors.

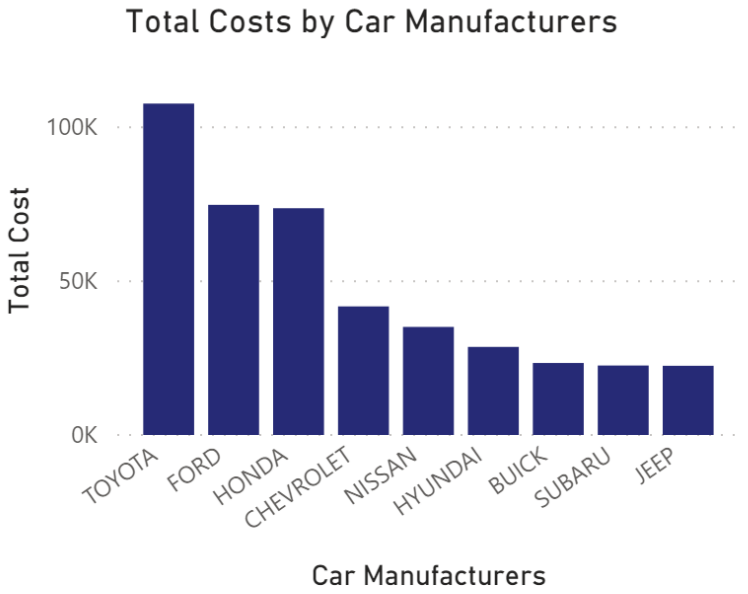
Member Engagement with Insurance

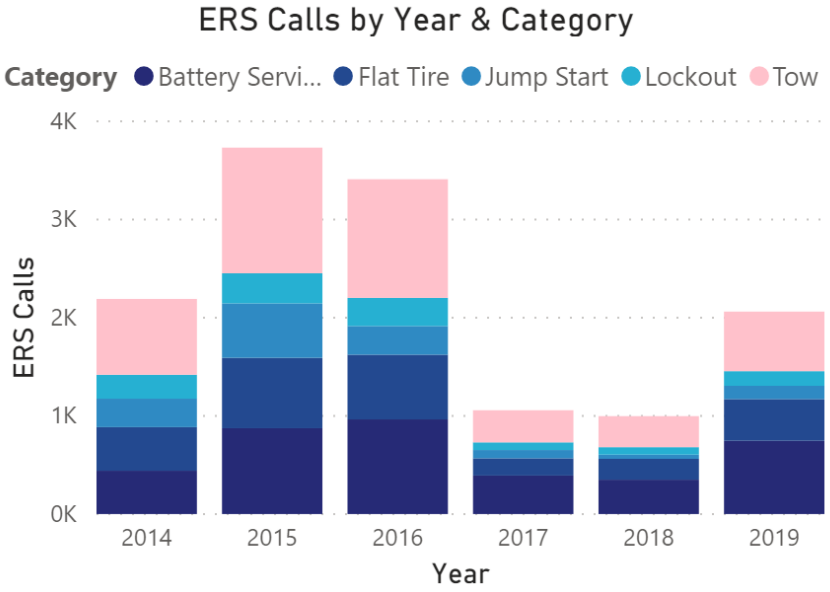


ERS Category and Costs



Exploratory Data Analysis





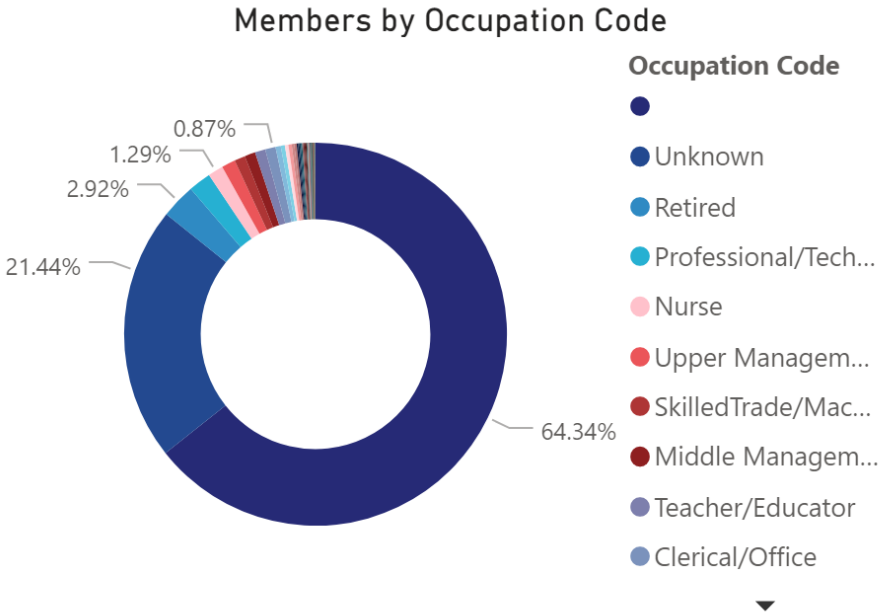
Findings

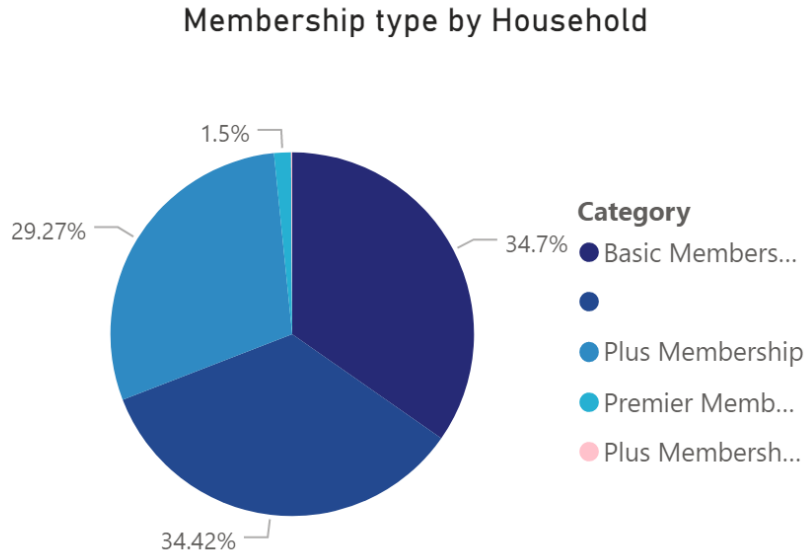
1.Toyota, Honda and Ford vehicles tend to need ERS more often.

2. Towing and Battery service calls are consistently high across 5 years.

3. 75% of occupation data is missing.

4. Financial and other insurance service products are not very popular among AAA NE members.





Insurance	By Household
N	4854
Y	1441
Total	**5241**

Credit Card	By Household
N	5099
Y	578
Total	**5241**

CMSI	By Household
N	5197
Y	368
Total	**5241**

Travel	By Household
N	5051
Y	731
Total	**5241**

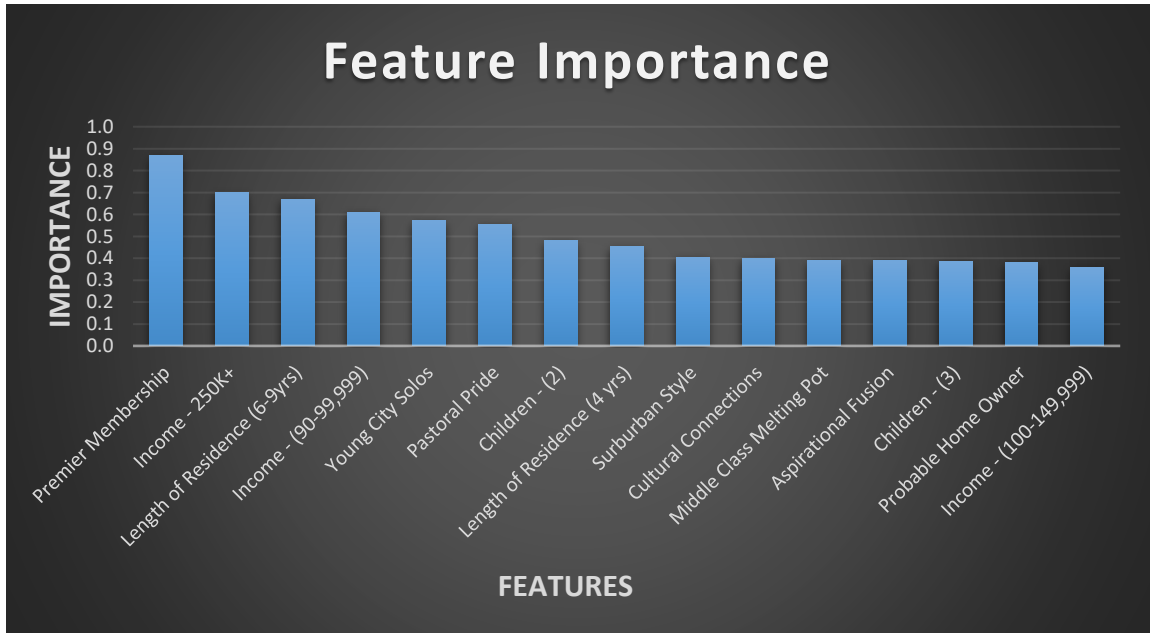
5. Predict Probability of Purchase

Model Setup

- Imbalanced dataset – Logistic Regression model built using 50/50 split.
- Probabilities calculated for Credit Card, Insurance, CMSI and Travel services.
- Validation data set was used to score resulting model.
- Predicted probabilities were used for Segmentation

Model Evaluation

The model accurately predicted a probability 60% of the time.



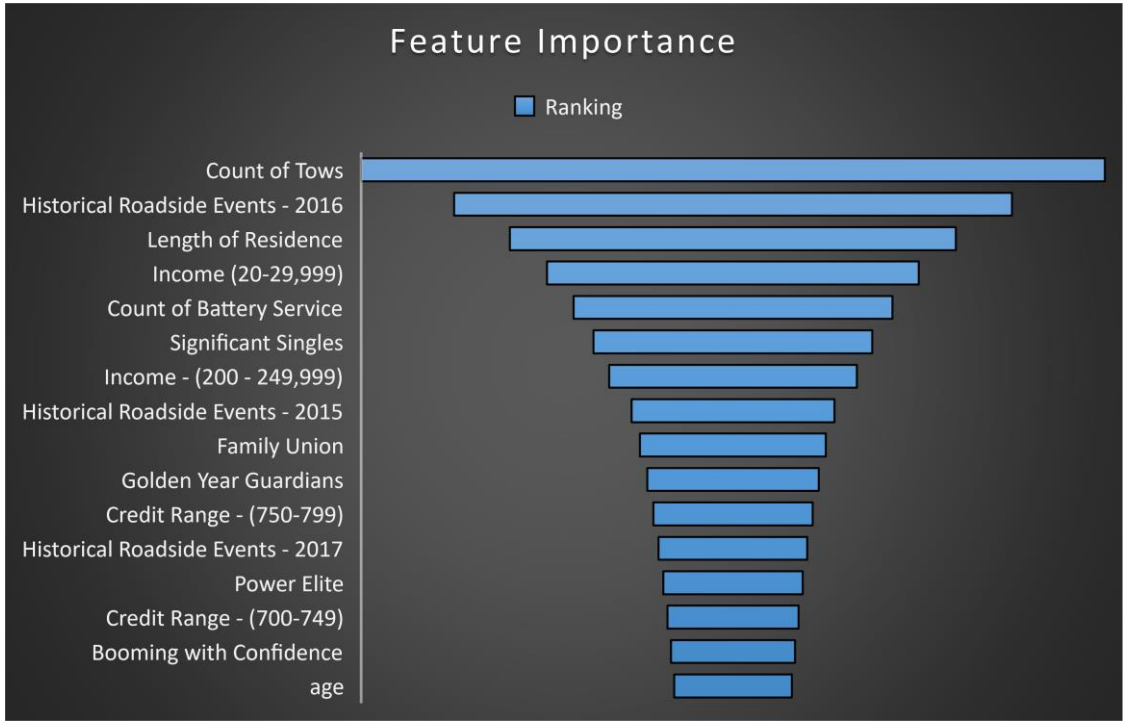
6. Predict Roadside Usage and Costs

Model Setup

- Data Cleansing, Feature Engineering applied and Correlation matrix used.
- Gradient Boosting Regressor Model Built and improved with RFCEV, GridSearch.
- Predicted values were compared with the original values. (Refer to Appendix)
- Other Regressor models were tested.
- Predicted Values were used for Segmentation.

Model Evaluation

Model Accuracy = 99.85%



RFECV - Recursive Feature Elimination and Cross Validation

7. Market Segmentation Overview

Cluster 0

★ **$52**

Per HH

- Ave Tenure – 35 years
- Earns $50K or more
- ERS – Tow & Battery Service
- May buy insurance services
- Basic Members – 50%
- ★ Autumn Years & Booming with Confidence

Cluster 1

$11

Per HH

- Ave Tenure – 38 years
- Earns $70K or more
- ERS –Battery & Tow Service
- Less probable to buy other services
- Mostly Basic Members
- Golden Guardians & Autumn Years

Cluster 2

$291

Per HH

- Ave Tenure – 26 years
- Earns $90K or more
- ERS – Tow & Battery Service
- Probable to buy insurance services
- Plus Member – 67%
- Autumn Years & Booming with Confidence

Cluster 3

$128

Per HH

- Ave Tenure – 30 years
- Earns 90K or more
- ERS –Tow & Battery Service
- Probable to buy other services
- Even spread - Basic & Plus Members
- Autumn Years & Booming with Confidence

★ *Annual Predicted Costs*

★ *Experian Mosaic Household Grouping*

Overall Clusters

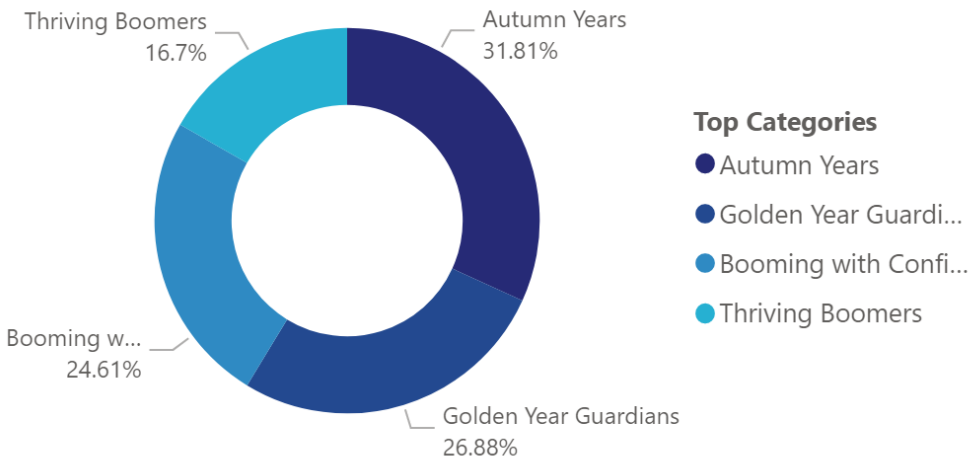
Total
5241
Household

Member Tenure
36
yrs per household

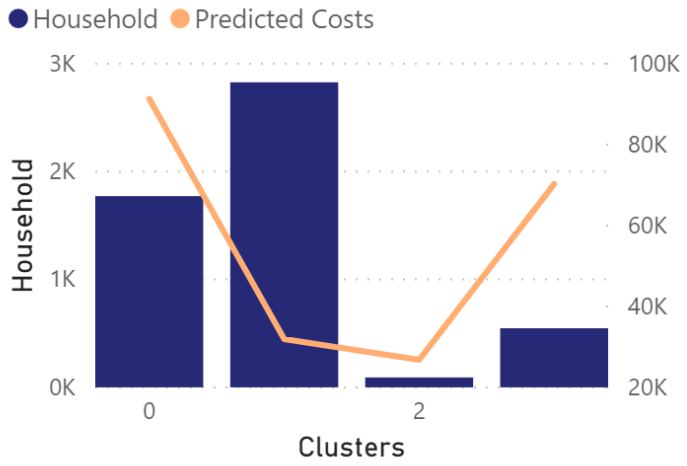
Average Age
72
yrs per household

Annual Predicted Cost
42
$ per household

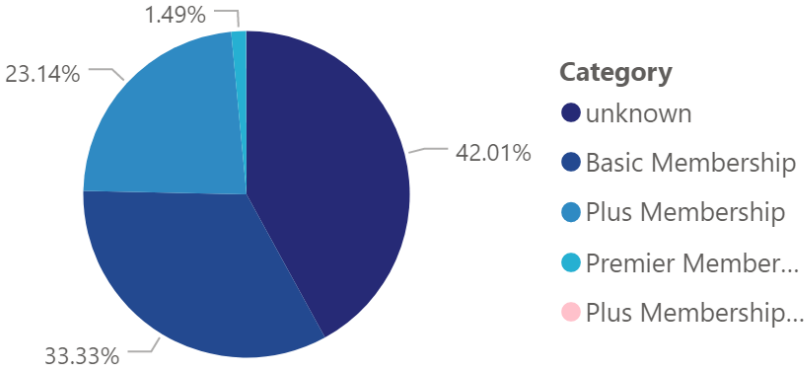
Experian Mosaic Grouping



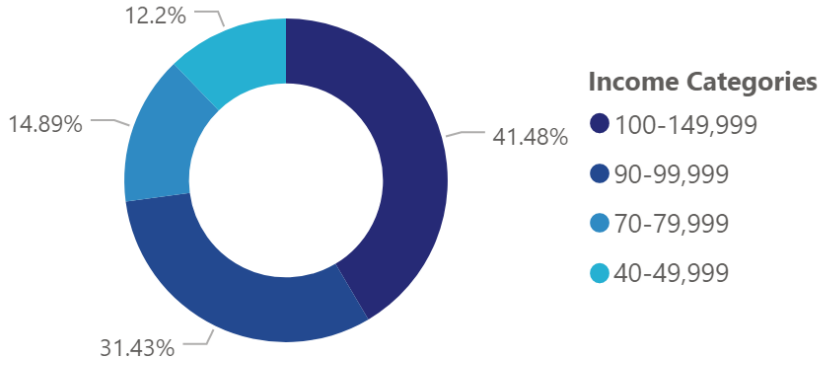
Cluster Trend - Predicted Costs vs No. of Household



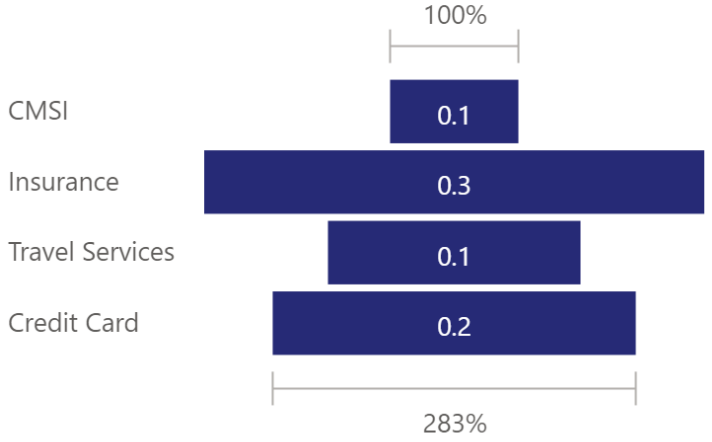
Household by Membership type



Household Income



Predicted Probability (Average)



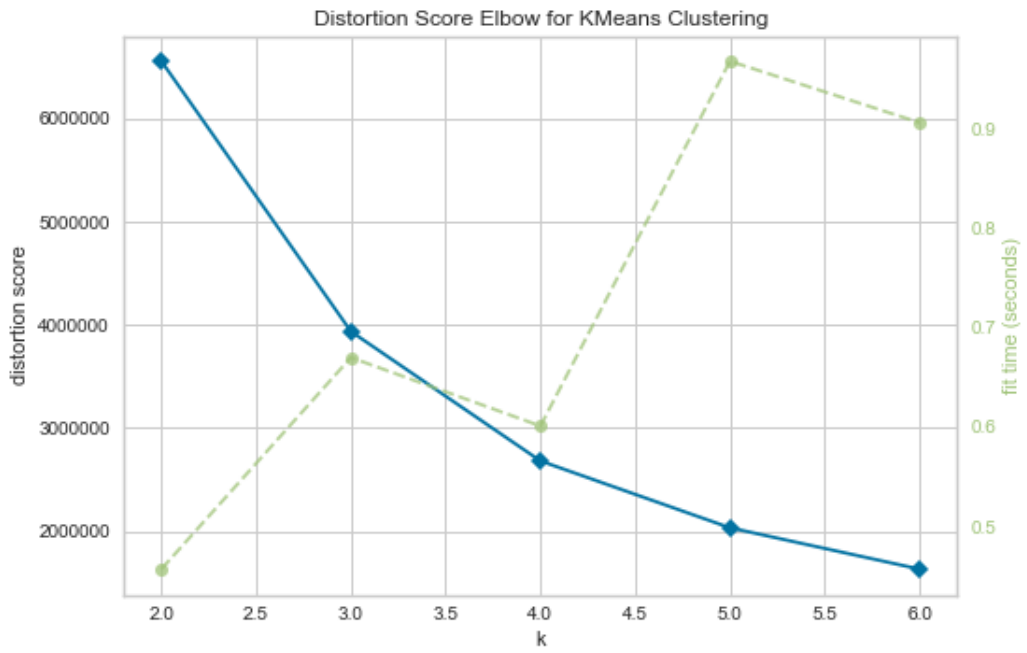
Market Segmentation

Model Setup

- Data Cleansing and Feature Engineering applied.
- Predicted roadside usage, costs and probability of purchase were used for clustering.
- K-Mean Clustering used to build the model.
- Elbow method used for selecting no. of clusters.
- Intercluster distances validated using Yellow bricks frameworks.

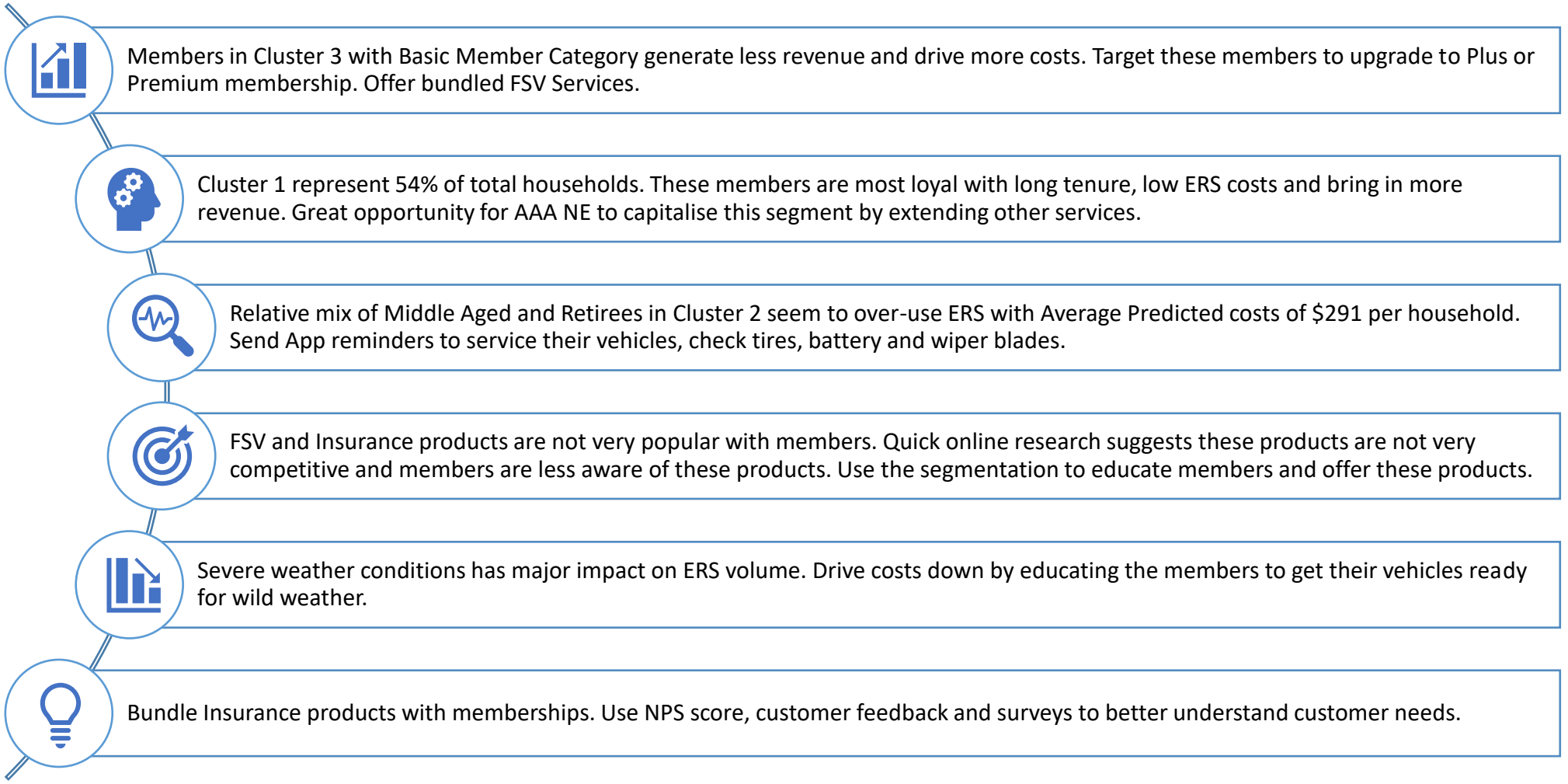
Cluster Selection

Clusters = 4



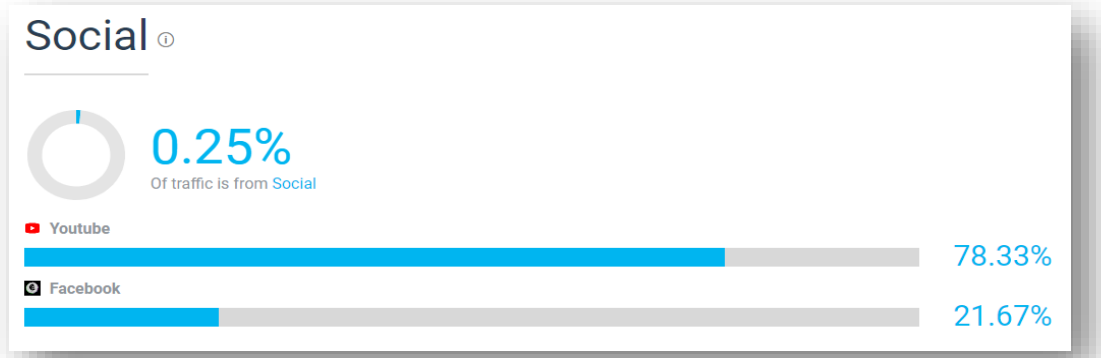
RFECV - Recursive Feature Elimination and Cross Validation

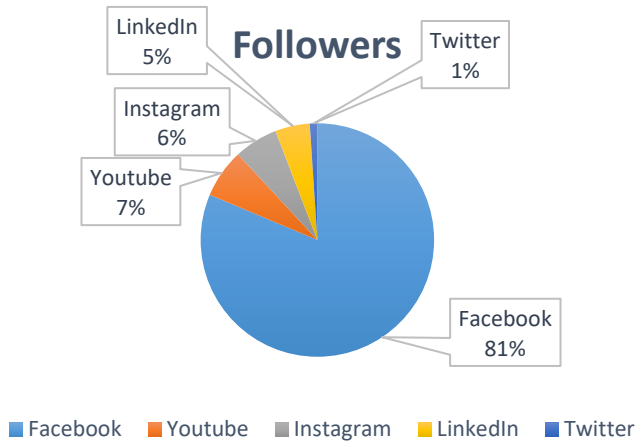
8. Key Recommendations

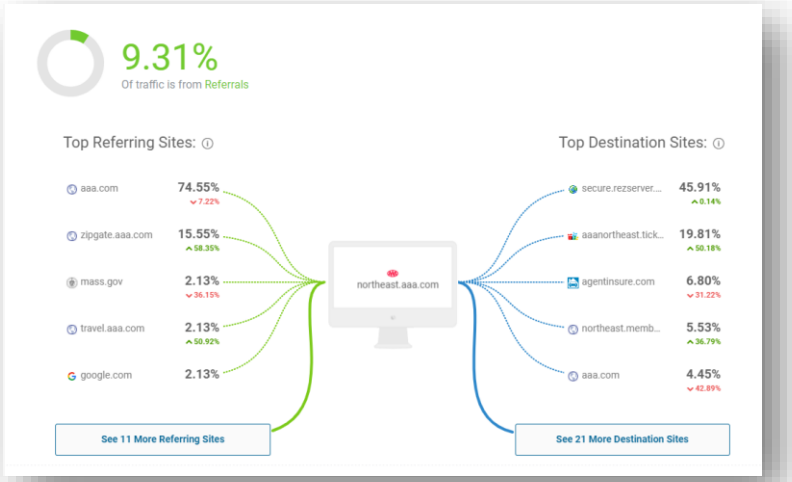


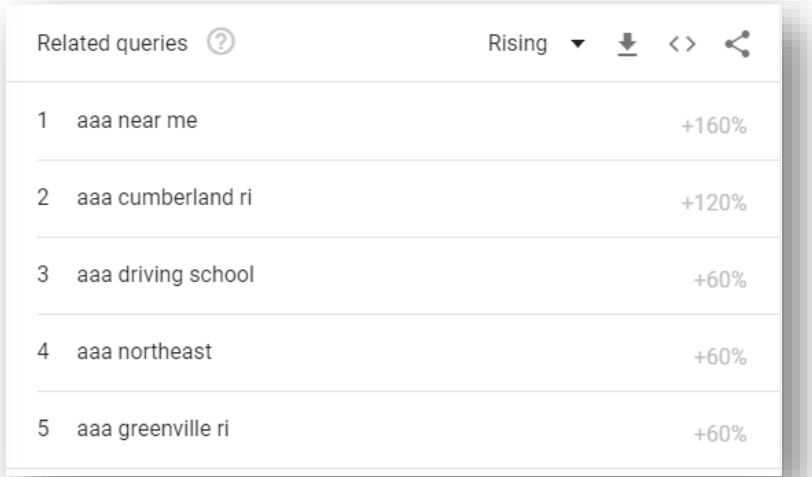
Improve Digital Engagement Strategy

- AAA NE may need to review their digital media strategy.
- Increase social media presence across various channels to increase engagement.
- AAA NE website has 48% bounce rate. Website forces viewers to enter zip code before one could access any information. Website should have common landing page and should subtly request zip code.
- Increase member satisfaction by improving digital experiences.









9. Appendix

- Market Segmentation – Individual Cluster Slides
- K-Means Cluster Plots.
- Gradient Boosting Model Plots.
- Logistic Regression Plots.

Cluster 0

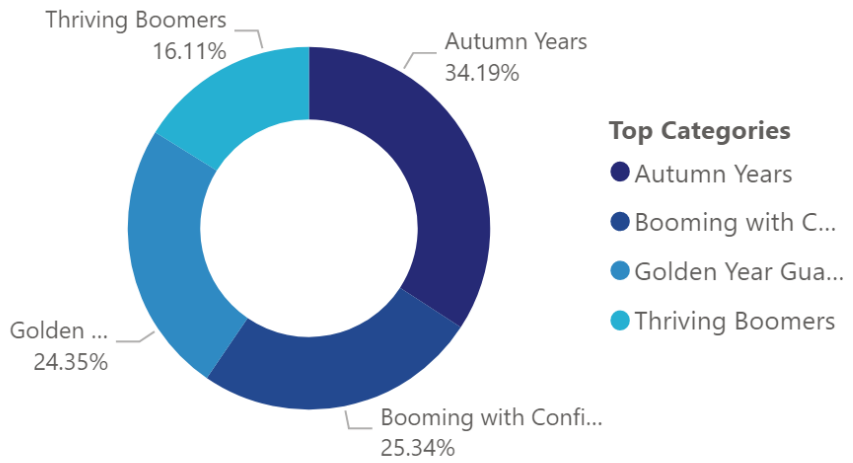
Total
1773
Household

Member Tenure
35
yrs per household

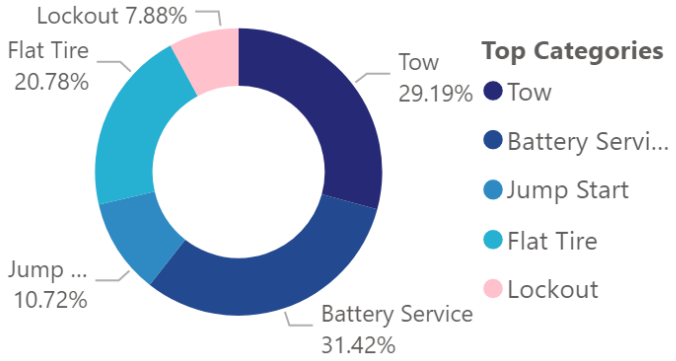
Average Age
71
yrs per household

Annual Predicted Cost
52
$ per household

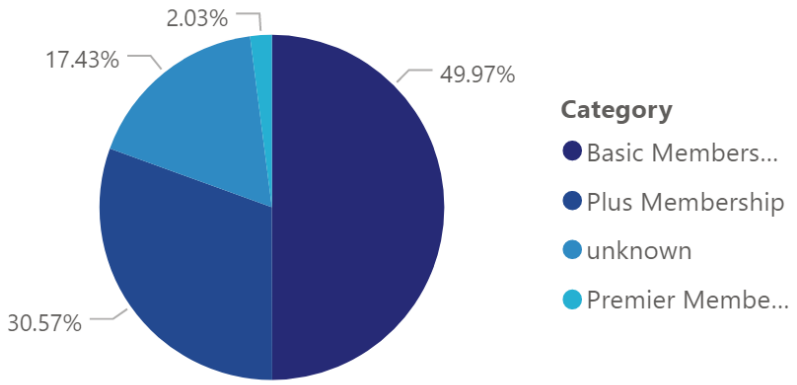
Experian Mosaic Grouping



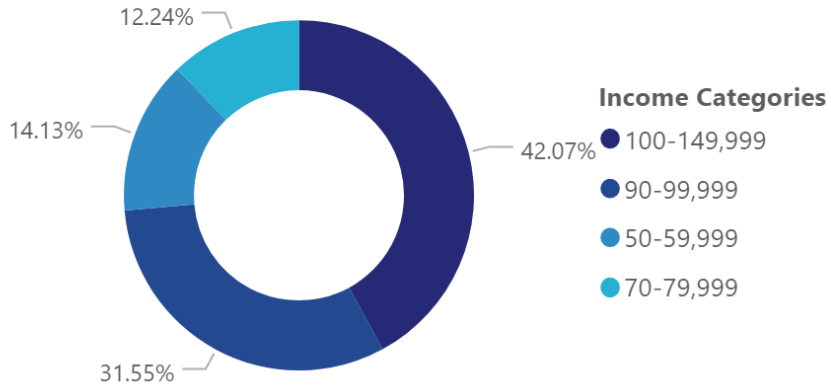
Type of ERS Needed



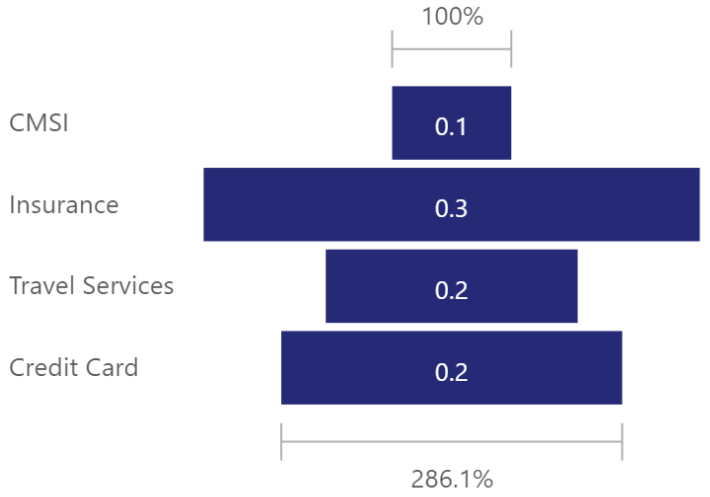
Household by Membership type



Household Income



Predicted Probability (Average)



Cluster 1

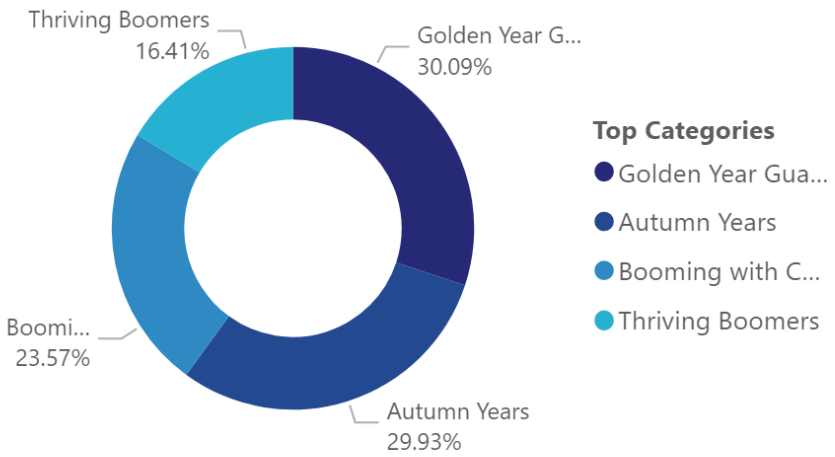
Total
2828
Household

Member Tenure
38
yrs per household

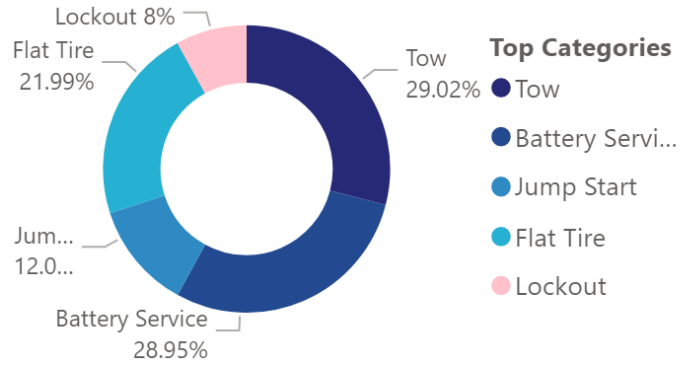
Average Age
74
yrs per household

Annual Predicted Cost
11
$ per household

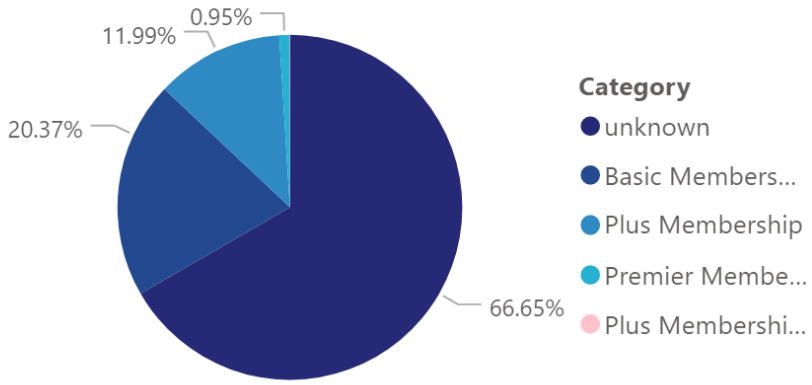
Experian Mosaic Grouping



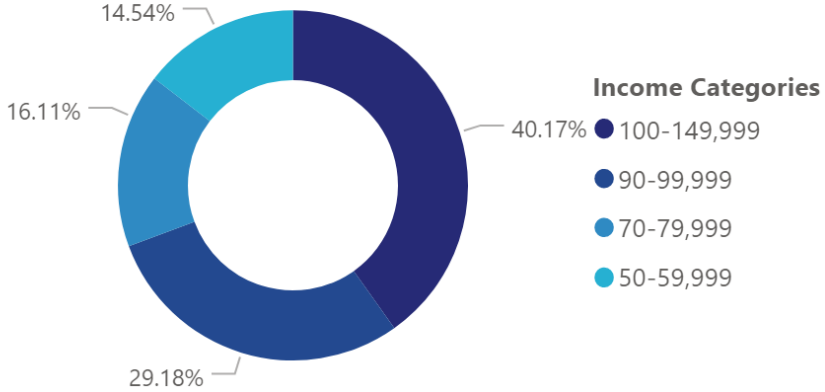
Type of ERS Needed



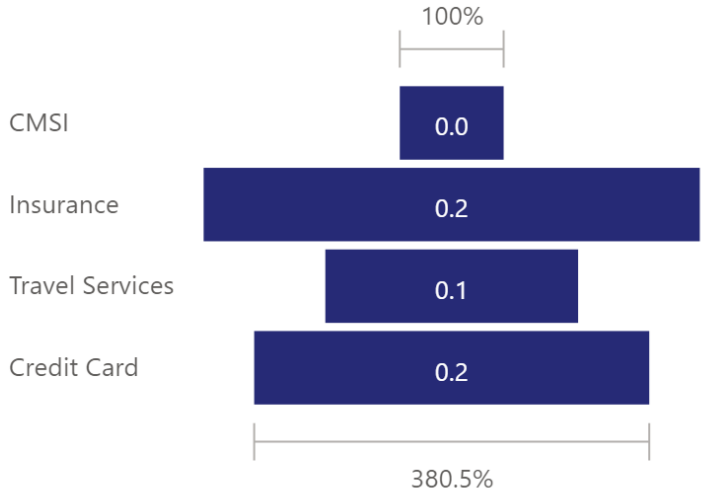
Household by Membership type



Household Income



Predicted Probability (Average)



Cluster 2

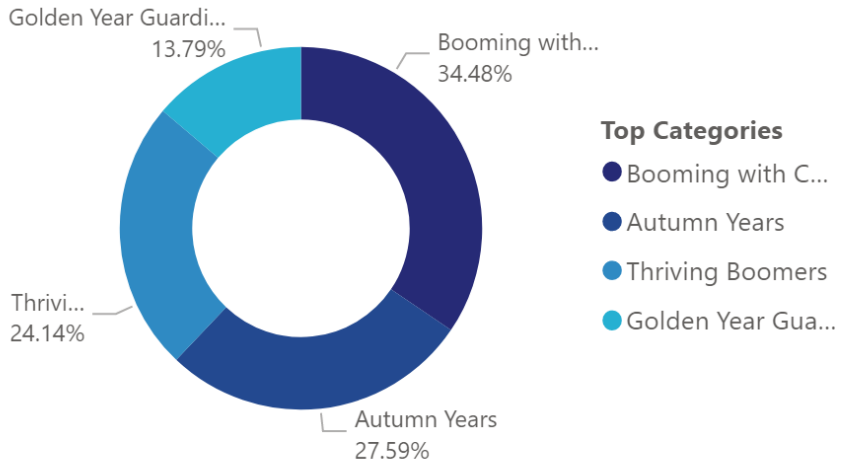
Total
92
Household

Member Tenure
26
yrs per household

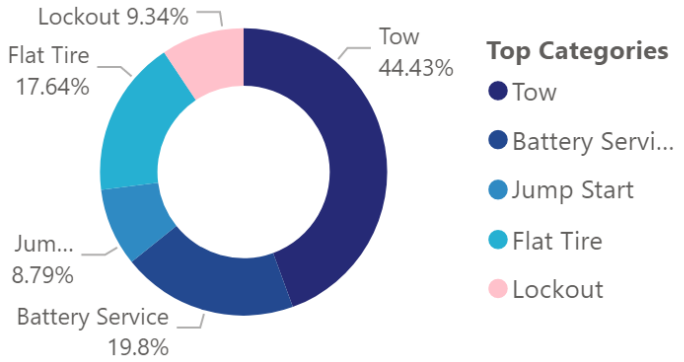
Average Age
61
yrs per household

Annual Predicted Cost
291
$ per household

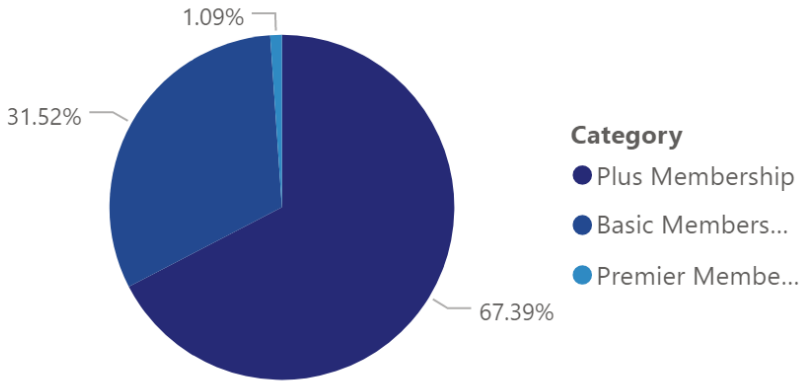
Experian Mosaic Grouping



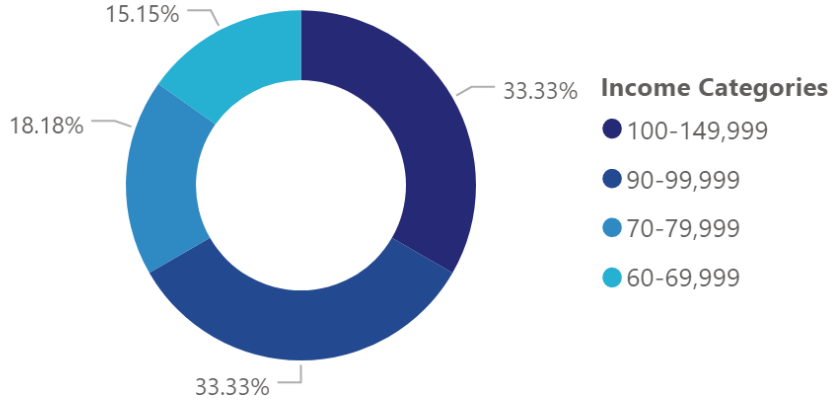
Type of ERS Needed



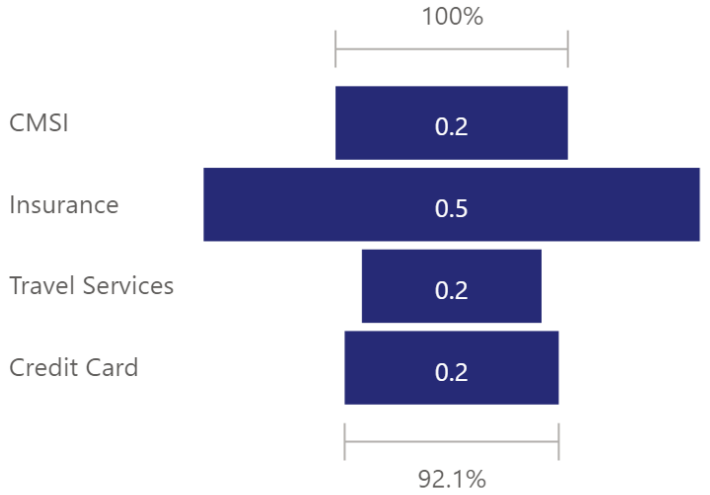
Household by Membership type



Household Income



Predicted Probability (Average)



Cluster 3

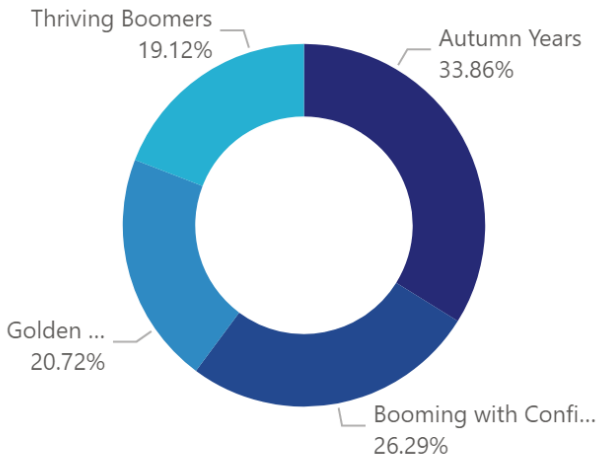
Total
548
Household

Member Tenure
30
yrs per household

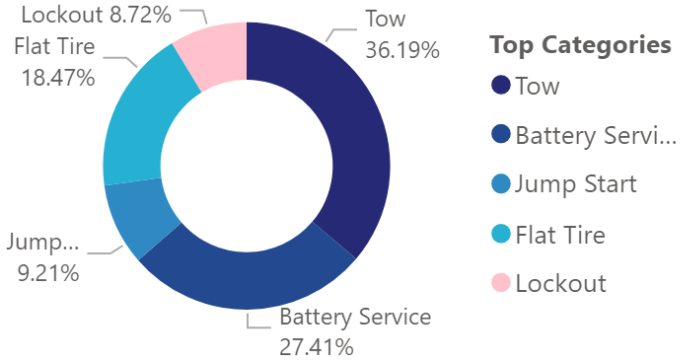
Average Age
68
yrs per household

Annual Predicted Cost
128
$ per household

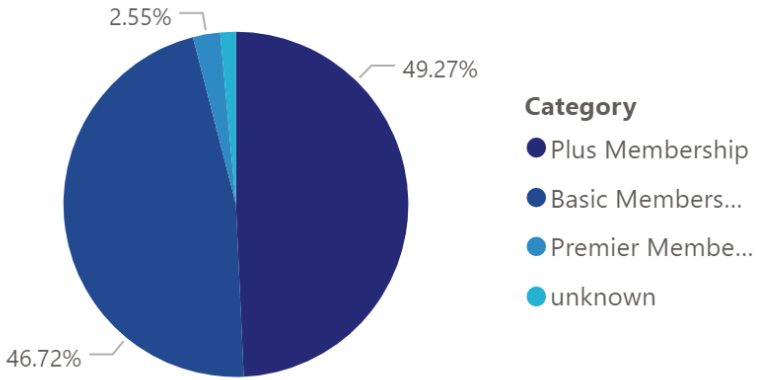
Experian Mosaic Grouping



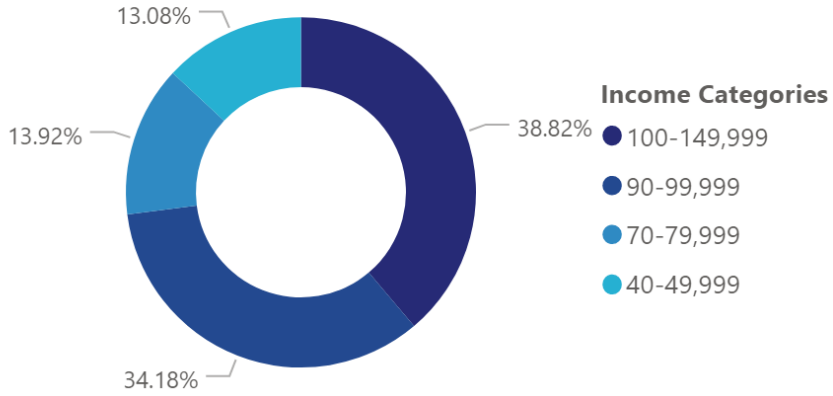
Type of ERS Needed



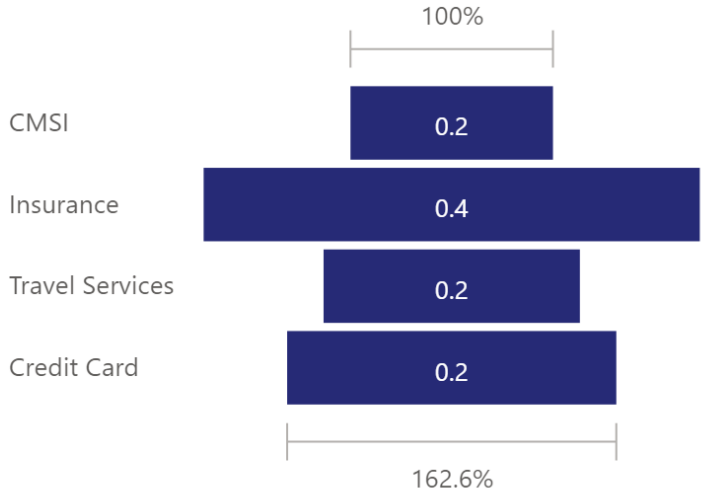
Household by Membership type



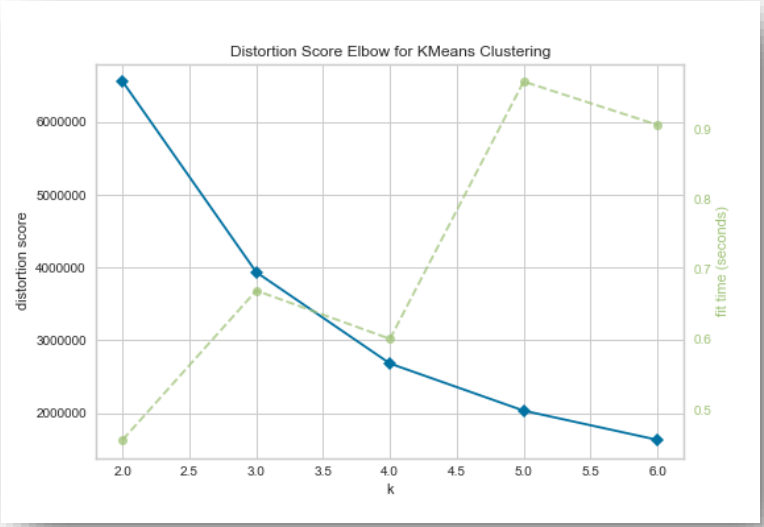
Household Income



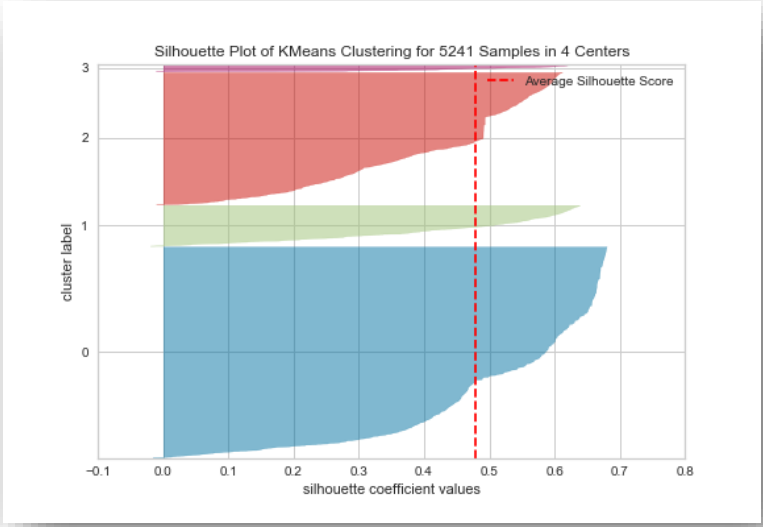
Predicted Probability (Average)



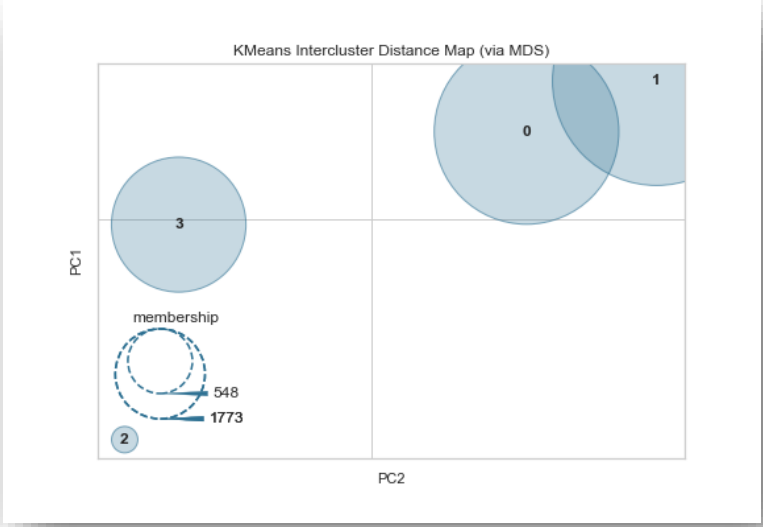
K-Means Cluster Plots



Elbow Method

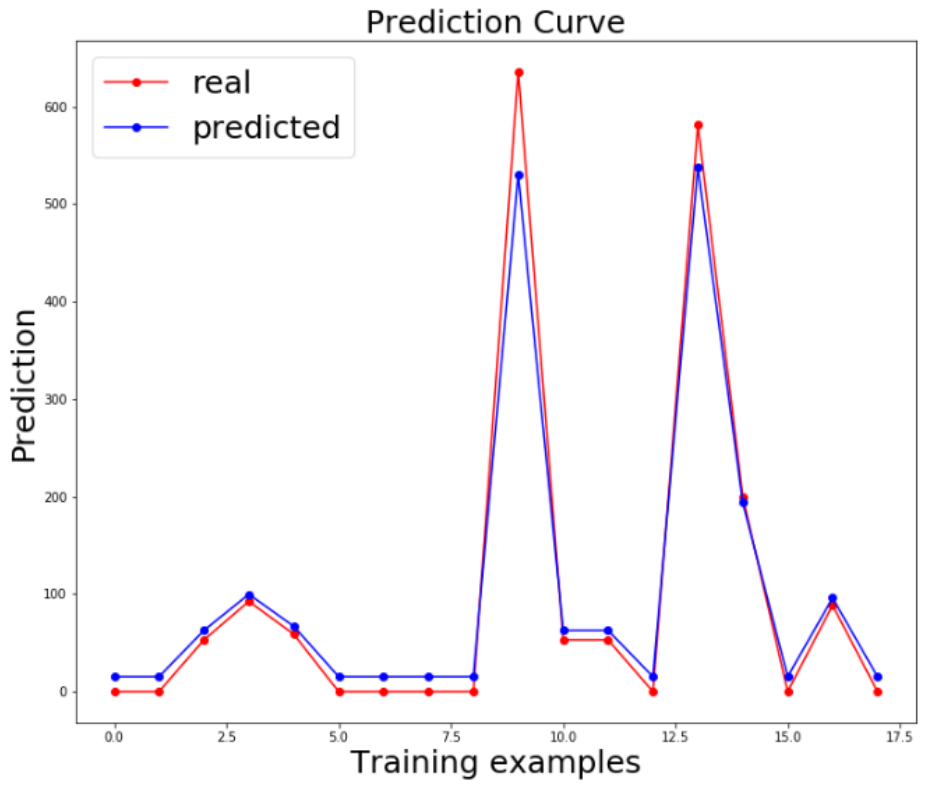


Silhouette Plot

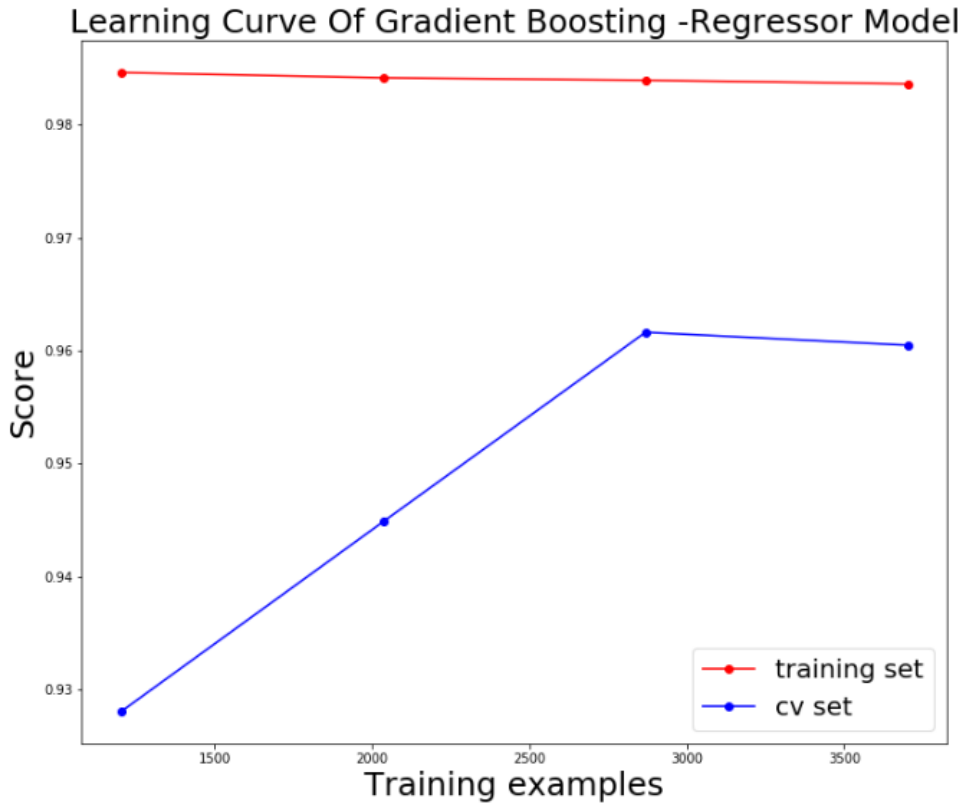


Inter Cluster Distance Map

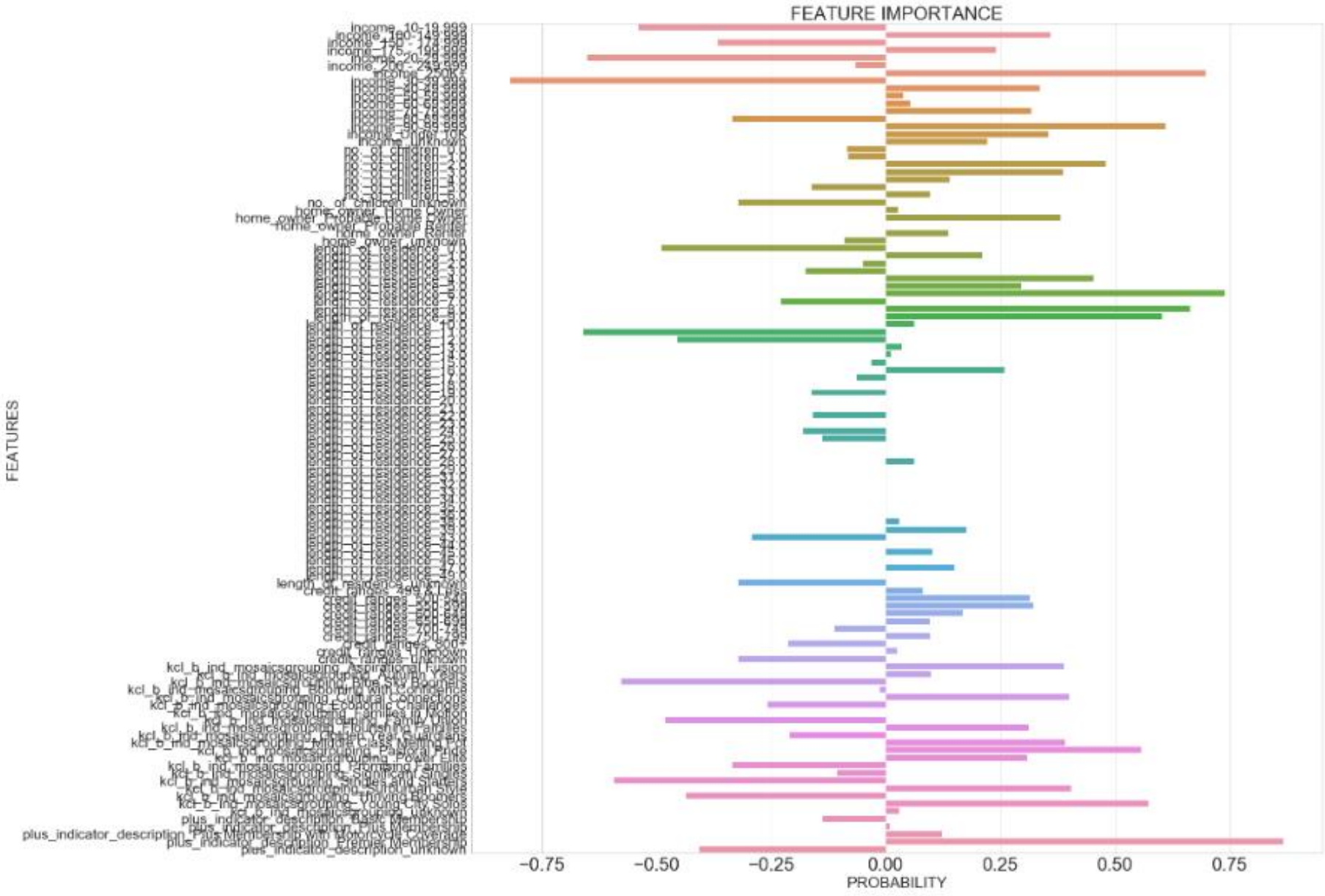
Gradient Boosting Regressor Plots



Prediction Curve



Learning Curve



Logistic Regression – Feature Importance Plot

10. References

Experian Mosaic Grouping

AAA Northeast

Gradient Boosting

K-Means Cluster Visualization

Average Weather in Providence

ERS – Emergency Roadside Services

RFECV - Recursive Feature Elimination and Cross Validation

THANK YOU