PGDMLAI Capstone Project - AAA Northeast Club

Background

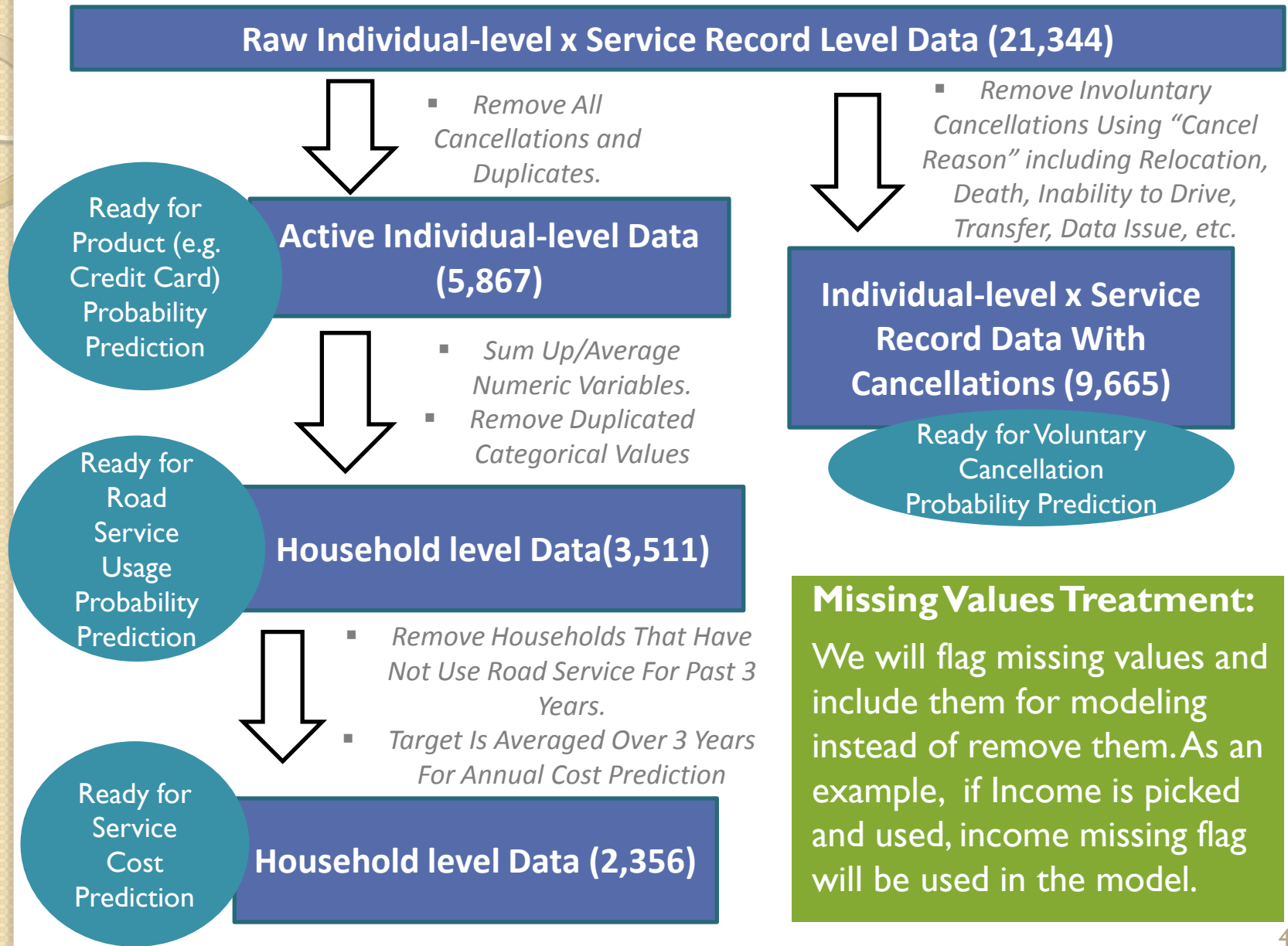- AAA Northeast Club Offer Roadside Services to all different level of members (Basic/Plus/Premier) with different level of benefits. Family member can also get benefits at reduced cost as associate members.

- Roadside Service is costly to AAA, especially towing.

- AAA also offers other products as well including insurance, credit card, travel ware, home equity/mortgage, and other financial services at competitive price.
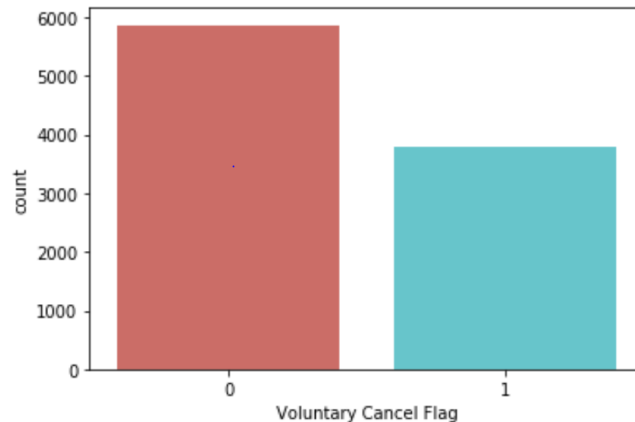
Objectives

Build segmentation of AAA member households:

- Better understand and anticipate the needs of members

- Customize communications and offering to various segments

- Expend more effort driving acquisition and renewal of desirable members

Data Preparation

Raw Individual-level x Service Record Level Data (21,344)

■ *Remove All Cancellations and Duplicates.*

Ready for Product (e.g. Credit Card) Probability Prediction

Active Individual-level Data (5,867)

■ *Remove Involuntary Cancellations Using "Cancel Reason" including Relocation, Death, Inability to Drive, Transfer, Data Issue, etc.*

Individual-level x Service Record Data With Cancellations (9,665)

Ready for Voluntary Cancellation Probability Prediction

■ *Sum Up/Average Numeric Variables.*
■ *Remove Duplicated Categorical Values*

Ready for Road Service Usage Probability Prediction

Household level Data(3,511)

■ *Remove Households That Have Not Use Road Service For Past 3 Years.*
■ *Target Is Averaged Over 3 Years For Annual Cost Prediction*

Ready for Service Cost Prediction

Household level Data (2,356)

Missing Values Treatment: We will flag missing values and include them for modeling instead of remove them. As an example, if Income is picked and used, income missing flag will be used in the model.
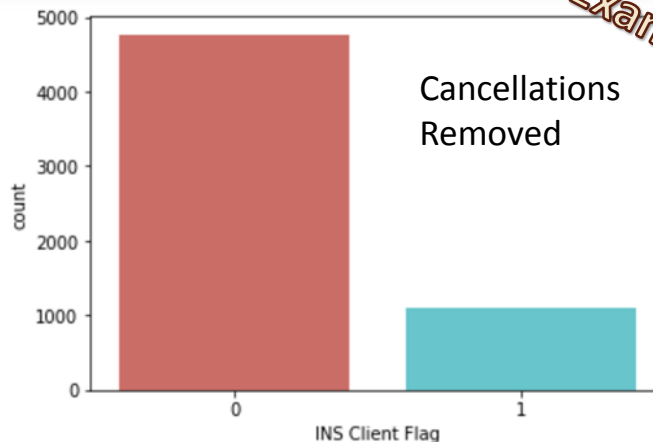
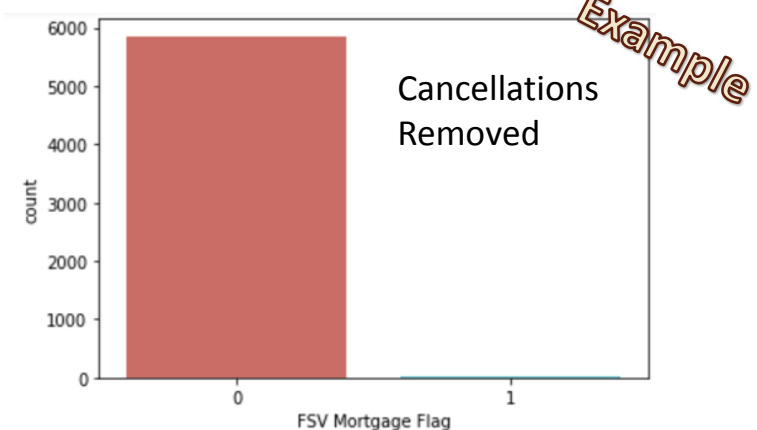# Exploratory Data Analysis (EDA): Individual-Level Data

High Percentage Of Voluntary Cancellations Occurred In Historical Data. It Would Add Values For Member Acquisition And Renewal.

Note: Involuntary Cancellations Uses "Cancel Reason" including Relocation, Death, Inability to Drive, Transfer, Data Issue, etc.

Cancellations Removed

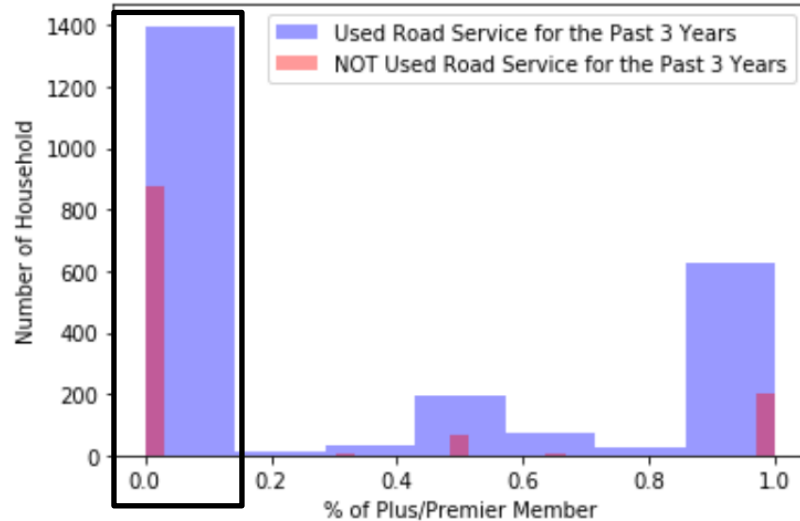Example

Cancellations Removed

Example

Reasonable Percentage Of Members Bought Insurance Products With AAA. Will Be Our Focus For Modeling.
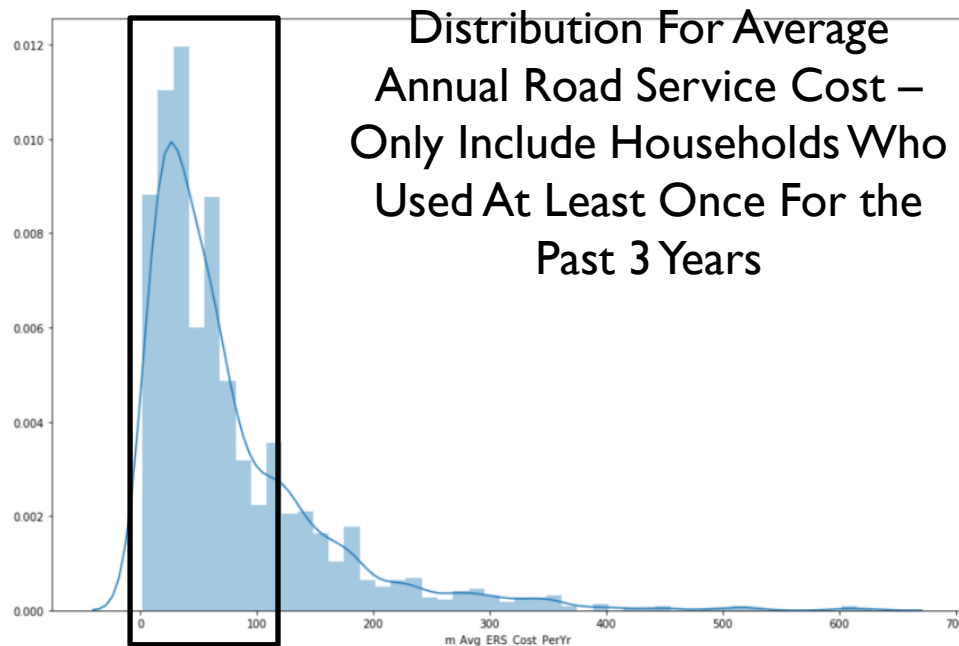
Near Zero Percentage Of Members Bought Mortgage With AAA. Model Will Have Problem Estimating On These.

Exploratory Data Analysis (EDA): Household-Level Data

> ➢ **Majority of the Household uses Road Service.**
>
> ➢ **Plus/Premier Members uses it more frequently than Basic Members.**

Distribution For Average Annual Road Service Cost – Only Include Households Who Used At Least Once For the Past 3 Years

> ➢ **Distribution for Road Service Cost is highly skewed.**
>
> ➢ **Majority of costs are Less than $100 for a year.**

Overview on Methodology & Process

- **Step 0: Sampling:** Random Sample data into training (60%) and Validation (40%) for each dataset/model.

- **Step 1: Feature Selection:** Use Logistic or Linear Regression to screen the variables based on the variable importance from the traditional statistical model.

- **Step 2: Estimation:** Build Multiple Machine Learning Model for each target such as Decision Tree, Random Forest, Extra Tree, Gradient Boosting.

- **Step 3: Model Selection:** Choose Champion Model for each target Based On accuracy on Validation Data which is not used for estimation.

- **Step 4: Predication:** Produce prediction using the champion model for validation data for each target.

- **Step 5: Data Aggregation and Summarization:** Summarize data into household-level and merge all the input data and predictions.

- **Step 6: Clustering:** Build K-Means Clustering model using data from above with K=3/4/5/6. K here means number of segments for household.

- **Step 7: Segment Selection:** Choosing Best Segmentation For Easy Implementation And Interpretation

*Note: More details information can be found in the Python codes.*

Results: Five Household Segments

- Household Are Segmented Into 5 Groups Using K-means Clustering Method. Estimated Probabilities Of Each Product, Probability Of Voluntary Cancellation, Probability Of Roadside Service Usage And Costs Are Used.

- Normalized Numbers Are Used In The Model To Ensure Each Field Is Weighted & Treated Equally.

Drivers Behind Different Segments

- **Main Drivers: Roadside Usage And Cost, Probability Of Cancellation.**
- **Minor Drivers: Probabilities Of Credit Card, Insurance & Travel ware.**
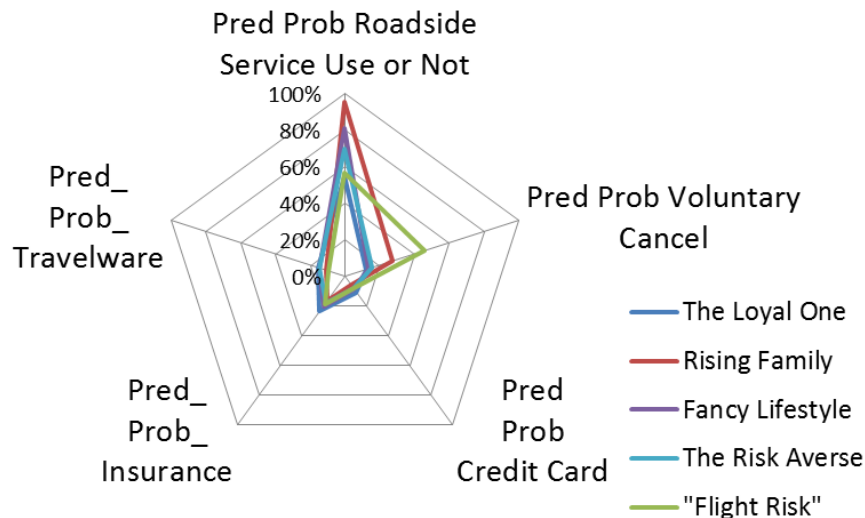
Household Segment	Number Of Household	% Count
The Loyal One	1,493	43%
Rising Family	422	12%
Fancy Lifestyle	723	21%
The Risk Averse	464	13%
"Flight Risk"	396	11%

Review Each Segment One By One In More Details For The Next 5 Pages.

Segment 1:The Loyal One

Highlight

➤ **Largest Segment** (43%).

➤ **Low Usage and Low Cost:** Lowest Roadside Usage Probability(55% For Past 3 Yrs) And Cost ($37 Per Yr).

➤ **Stable Revenue:** Lowest Probability Of Voluntary Cancellation (12%) With Highest Tenure Years (Avg. 41 Years For Nonmissing).

➤ **High Engagement:** Highest Probability To Buy Other Aaa Products Including Credit Card (11%), Insurance (23%) And Travel Ware (15%).

➤ **High Revenue Potential:** On Average 1.1 AAA Members In The Household .Mostly Primary Membership (92% On Avg.).

➤ **Stable Financial Situation:** Relatively Low Income (Avg. $82k For Nonmissing). High Homeownership (74%). Low Amount of Car Ownership (Avg. 1.3 Cars For Nonmissing over past 6 years).

Recommendation

➤ **Promote and Offer Discounted Price To Other Family Members For Primary, Associate And Gift Membership.**

➤ **Promote and Offer Discounted Price For Plus & Premier Membership.**

➤ **Promote On AAA Products Regularly.**

Segment 2: Rising Family

Highlight

➢ **Small Segment** (14%).

➢ **High Usage and High Cost:** Highest Roadside Usage Probability(95% For Past 3 Yrs) And Cost ($101 Per Yr).

➢ **Low Stability:** Medium Probability Of Voluntary Cancellation (27%) With Lowest Tenure Years (30 Years For Nonmissing).

➢ **Low Engagement:** Lowest Probability To Buy Other AAA Products Including Credit Card (5%), Insurance (16%) And Travel Ware (11%).

➢ **High Revenue:** Three Members In The Household On Avg.

➢ **High Revenue Potential:** Lowest Member Age (Avg. 60 Years Old For Nonmissing). Number Of Childs Is Roughly 3 Per Household On Avg.

➢ **Improving Financial Situation:** Relatively High Income (Avg. $110k For Nonmissing). Relatively Low Homeownerships (63% for nonmissing). High Amount Of Car Ownership (Avg. 5.2 Cars For Nonmissing Over Past 6 Yrs).

Recommendation

➢ **Promote Plus & Premier Membership To Increase Profit Given High Cost.**

➢ **Survey On Their Satisfactions For AAA Roadside Services. Aim To Lower Cancellation And Build Long-term Membership.**

➢ **Survey On Their Considerations For AAA Products. Understand Their Needs.**

Segment 3: Fancy Lifestyle

Highlight

- **Second Largest Segment** (23%).

- **Medium Usage and Medium Cost:** Highest Roadside Usage Probability(81% For Past 3 Yrs) And Cost ($59 Per Yr).

- **Stable Revenue:** Low Probability Of Voluntary Cancellation (13%) With Avg. Tenure Years (36 Years For Nonmissing).

- **Mixed Engagement: High Probability To Buy Travel Ware (15%**) But Average On Others.

- **Medium Revenue Potential:** 1.7 Members In The Household On Avg. And Number Of Childs Are 1.4 On Avg.(For Nonmissing).

- **Good Financial Situation:** Relatively High Income (Avg. $110k For Nonmissing). High Homeownship (75% For Nonmissing). High Amount Of Car Ownership (Avg. 2.7 Cars For Nonmissing Over Past 6 Yrs).

Recommendation

- **Offer More Benefits And Discounts With Travel Ware And Other Exclusive Benefits Related To Hotels, Flights, Restaurants, Resorts, Private Events, etc.**

- **Survey On Their Considerations For Other AAA Products. Understand Their Needs.**

Segment 4: The Risk Averse

Highlight

➢ **Small Segment** (15%).

➢ **Medium Usage and Low Cost: High Roadside Usage Probability(70% For Past 3 Yrs) And Low Cost For The Usage ($29 Per Yr).**

➢ **Stable Revenue:** Low Probability Of Voluntary Cancellation (15%) With Avg. Tenure Years (37 Years For Nonmissing).

➢ **Mixed Engagement: High Probability To Buy Travel Ware (15%**) But Average On Others.Lowest Plus & Premier Membership (29% On Avg. For Nonmissing).

➢ **Medium Revenue Potential:** 2.3 Members In The Household On Avg. And Number Of Childs Are 1.8 On Avg.(For Nonmissing).

➢ **Good Financial Situation:** Relatively High Income (Avg. $110k For Nonmissing). High Homeownship (76% For Nonmissing). High Amount Of Car Ownership (Avg. 2.1 Cars For Nonmissing Over Past 6 Yrs).

Recommendation

➢ **Offer More Benefits And Discounts With Travel Ware And Other Exclusive Benefits Related To Hotels, Flights, Restaurants, Resorts, Private Events, etc.**

➢ **Promote Plus & Premier Membership.**

➢ **Survey On Their Considerations For Other AAA Products. Understand Their Needs.**

Segment 5:"Flight Risk"

Highlight

- **Small Segment** (13%).

- **Low Usage and Low Cost: High Roadside Usage Probability (57% For Past 3 Yrs) And Low Cost For The Usage ($35 Per Yr).**

- **Unstable Revenue:** Highest Probability Of Voluntary Cancellation (45%).

- **Low Engagement:** High Probability To Buy Other Aaa Products Including Credit Card (7%), Insurance (19%) And Travel Ware (10%).

- **Huge Missing Information: 45% Missing On Member Profiles Such As Number Of Childs, Credit, Income And 84% Missing On Occupations.**

- **Relatively Low Revenue Potential:** Highest Member Age (Avg. 82 Years Old). 1.4 Members In The Household On Avg.. Low Number Of Childs (0.7 On Avg. For Nonmissing).

- **Decent Financial Situation:** Relatively Low Income ($ 82k On Avg. For Nonmissing) And Low Car Ownership (1.3 On Avg. For Nonmissing).

Recommendation

- **Offer Discounted Price To Renew. Aim To Lower Cancellation Rate And Build Long-term Membership. Survey On Their Satisfactions For AAA Roadside Services**

- **Survey On Their Information To Better Understand Them And Tailor The Actions.**

Variable Importance & Revenue

Another View To Look At Important Variables And Their Relationship With Revenue:

➢ **Higher The Better: Tenure Years, Member Ages, Length Of Residence, Primary Member Flag, Higher Education Flag, Homeowner Flag, Auto Renew Flag, Etc.**

➢ **Lower The Better: Plus/Premier Members, New Mover Flag, Income, Credit Score, European Brand Cars, Luxury Brand Cars, Etc.**

Key Models	Tenure Year	Member Age	Length of Residence	Primary Member Flag	Retired Flag	College or Higher Education	Mail Respond Yes	Homeowner Flag	Renter flag	Renter flag	Auto Renew Flag
Probability of Credit Card	+	+		+		+					+
Probability of Insurance	-			+					+	+	+
Probability of Travel Ware	+	+				+					
Probability of Voluntary Cancel	-	+	+	+	+	+		+			
Probability of Road Service Usage	+	-		+							
Road Service Cost	+		-	-				+	+		

Key Models	Plus/ Premier Member Flag	New Mover Flag	Income	Credit Score	European Car Manufacturer	Japan Car Manufacturer	American& European Luxury Car Manufacturer
Probability of Credit Card			-	-			
Probability of Insurance			-	-			
Probability of Travel Ware							
Probability of Voluntary Cancel			-	+			
Probability of Road Service Usage	-				-	-	
Road Service Cost	-	-			-		-

Positive Sign Means Higher Variables' Value Correlates With Higher Revenue. Negative Sign Means The Opposite.

Note: Raw Relationships Between Models And Variables Can Be Found In The Appendix.

14

Recap

- **Five Distinct Type Of Households:** 1:the Loyal One**; 2:**rising Family; 3: Fancy Lifestyle; 4: The Risk Averse; 5: "Flight Risk". Tailoring Communications And Offerings Based On Various Families' Needs Is Critical.

- **Revenue From Membership**: Key Variables And Its Relationships With Revenue Provides Information Who We Want To Acquire And/Or Renew With The Most Efforts And Time.

- **More Information Needed For Better Accuracy:** There Are Roughly 40% Data Missing For Characteristics On Individual-level Data. 33% Data are Missing for Cars Brands and Models and membership type (Plus/Premier). 100% are missing for cars' age and mileage.
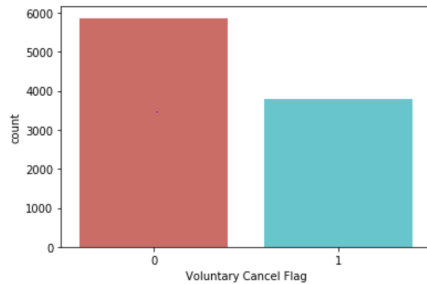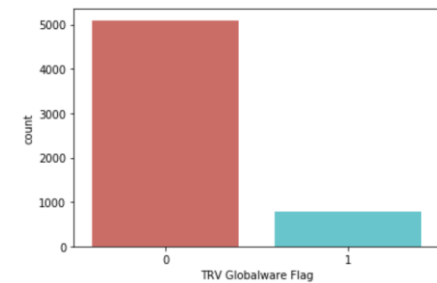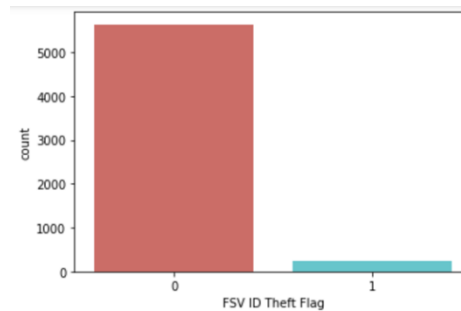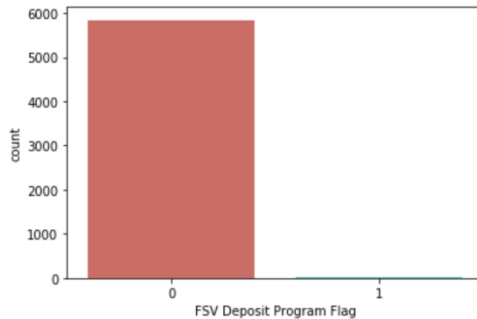
Appendix

EDA – All Targets

17
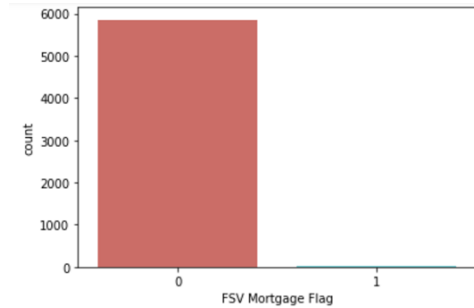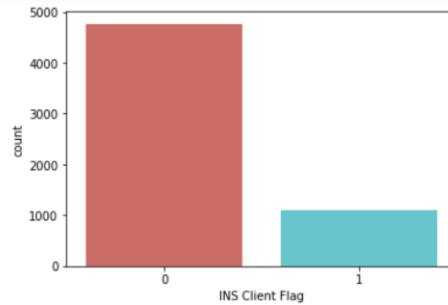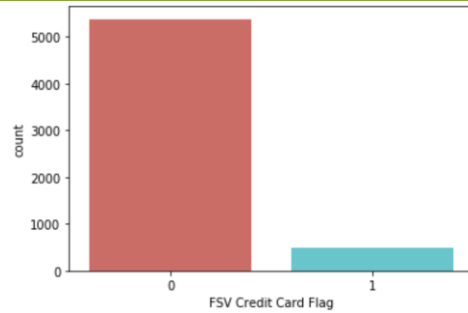
Correlation – Roadside Usage Count vs Cost Per Year

- Very High Linear Correlation: Pearsons Correlation = 0.94. It means
- Decided To Use Dollar Cost Only For Modeling.
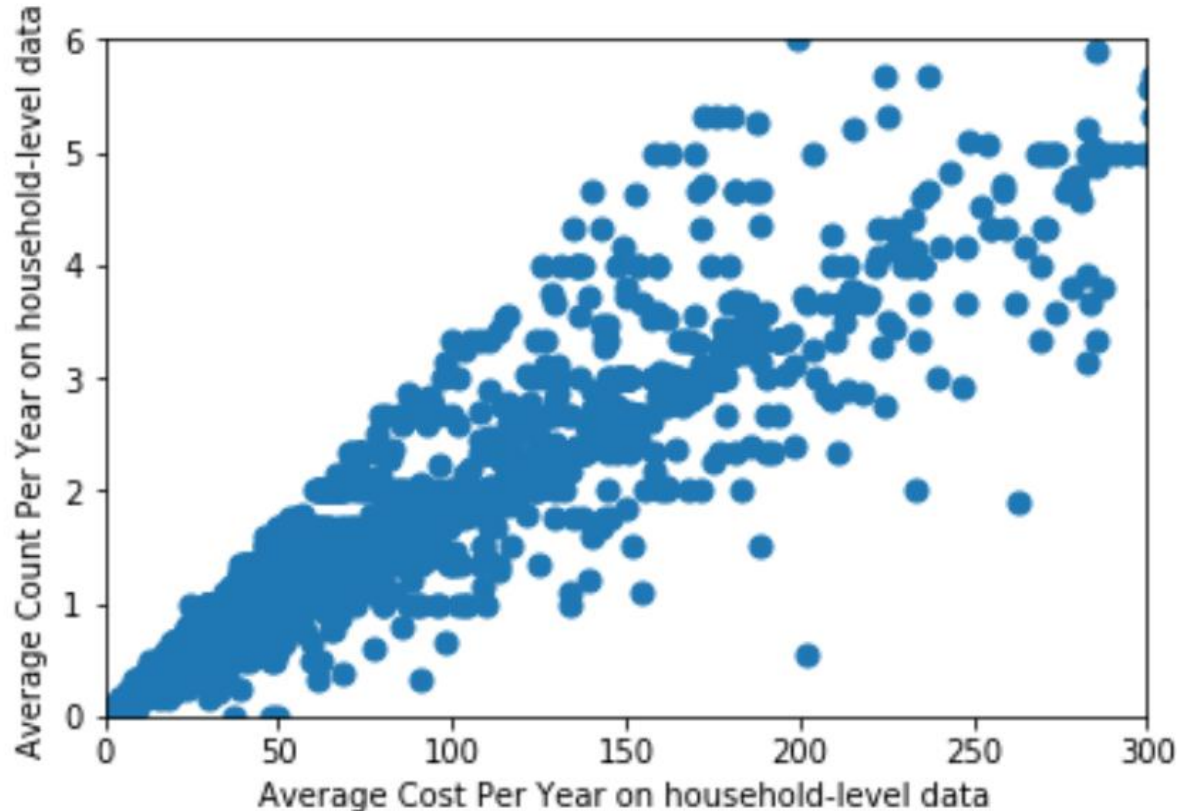
Data Table by Segments: Prediction and Key Characteristics.

Model Estimation

Household Segment	Total Count Household	Pred Service Cost Per Year	Pred Prob Roadside Service Use or Not	Pred Prob Voluntary Cancel	Pred Prob Credit Card	Pred_Prob_Insurance	Pred_Prob_Travelware	Pred Prob CMSI1	Pred Prob Deposit1	Pred Prob Home Equity1	Pred Prob Mortgage1	Pred Prob IDtheft1
The Loyal One	1,493	$ 37	55%	12%	11%	23%	15%	4%	0.4%	0.0%	0.2%	5%
Rising Family	422	$ 101	95%	27%	5%	16%	11%	7%	0.4%	0.1%	0.2%	3%
Fancy Lifestyle	723	$ 59	81%	13%	8%	20%	15%	5%	0.4%	0.1%	0.2%	4%
The Risk Averse	464	$ 29	70%	15%	8%	18%	15%	5%	0.4%	0.1%	0.2%	4%
"Flight Risk"	396	$ 35	57%	45%	7%	19%	10%	3%	0.4%	0.1%	0.2%	4%

Household Characteristics

Household Segment	Number Of Household	% Count	Avg. Number of Member in the household	Avg. Member age (For nonmissing)	Number of Childs (For nonmissing)	% Homeowner (For nonmissing)	% Single Family Dwelling (For Nonmissing)	Income (For nonmissing)	Number of Cars (Limited)	Tenure Year (For nonmissing)	% Primary Member	% Plus&Premier Membership
The Loyal One	1,493	48%	1.1	76	0.6	74%	62%	82,018	1.3	41	92%	48%
Rising Family	422	14%	3.1	60	2.8	63%	70%	110,371	5.2	30	37%	44%
Fancy Lifestyle	723	23%	1.7	71	1.4	75%	68%	108,757	2.7	36	66%	48%
The Risk Averse	464	15%	2.3	70	1.8	76%	75%	108,829	2.1	37	46%	29%
"Flight Risk"	396	13%	1.4	82	0.7	54%	40%	82,018	1.3	36	83%	42%
Total	3,102	100%										

Variable Importance By Models

Positive Sign Means Higher Variables' Value Correlates With Higher Target Value (Probability & Numbers). Negative Sign Means The Opposite.

	Tenure Year	Member Age	Length of Residence	Income	Credit Score	Primary Member Flag	New Mover Flag	Retired Flag	Higher Eduation	Mail Respond Yes	Homeowner Flag	Renter flag	Auto renew Flag	Plus/Primier Member Flag	European Car Manufacturer	Japan Car Manufacturer	American& Europrean Luxury Car Manufacturer
Probability of Credit Card	+	+		-	-	+			+				+				
Probability of Insurance	-			-	-	+						+	+				
Probability of Travel Ware	+	+							+								
Probability of Voluntary Cancel	+	-	-	-		-	+	-	-				-				
Probability of Road Service Usage	-	+				-								+	+	+	
Road Service Cost	-		+			+	+			-	-			+	+		+