



# Churn at Car Dealerships

---

Khalid, Rachel & Sherry  
April 27<sup>th</sup>, 2020

# Executive Summary

- Dealer-FX Car Dealerships are losing revenue opportunities via average annual churn of 40% in their Service Departments
- Churn Prediction model was developed to quantify customer loss probability
  - Customers ranked based on predicted risk
  - Highest risk group predicted to be 2 times more likely to churn than the average rate
  - High agreement between predicted customer risk and actual customer churn
- Model to guide pro-active targeting of high risk customers through marketing campaigns
  - Customer scoring mechanism used to segment and profile top risk customers based characteristics
  - Car Dealerships with riskiest customers can be identified for potential marketing campaigns

# Presentation Roadmap

## Analysis Framework



Dealer Segmentation  
& Variable Definitions

Churn Prediction Model



Attrition Reasons  
& Customer Story



Deliverables & Future Work



# Business & Analytical Objectives



**Tailor a Solution for Car Dealerships to Reduce Customer Churn in their Service Department, via:**

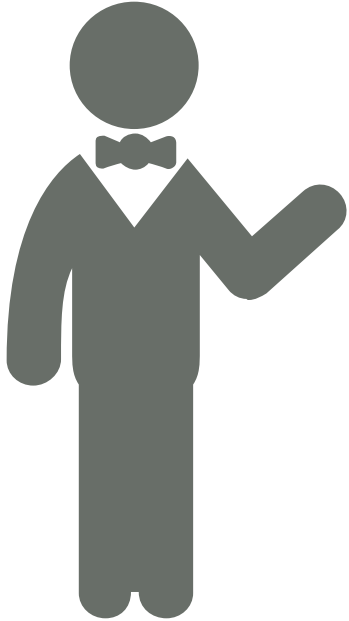
**Identifying:** Customer Service Visits from 2016 - 2018

**Predicting:** Customer Churn Probability for 2019

**Evaluating:** Customer Attrition Reasons

**Implementing:** Customized Dealer Solution

# Project Setting



- Data Information:
  - Analyzed customer repair orders (RO) from the period of November 2016 – November 2019
- Customer Definition:
  - Unique Customer = Repair Order Customer ID + VIN
- Customer Churn Label:
  - Retained Customer: Showed up at least once in two consecutive years
  - Lost Customer: Showed up at least once between 2016 & 2018 and did not show up in the following year
  - New Customer: Showed up at least once in 2019
- Customer Status:
  - Customer Churn label applied at end of Year 3 (2019)
  - Observation Period: Years 1 & 2 (2016 –2018)
  - Prediction Period: Year 3 (2019)

# Customer Service Visits



**First Step**



Segmenting all dealers in the USA



**Second Step**



Grouping dealers based on demographics

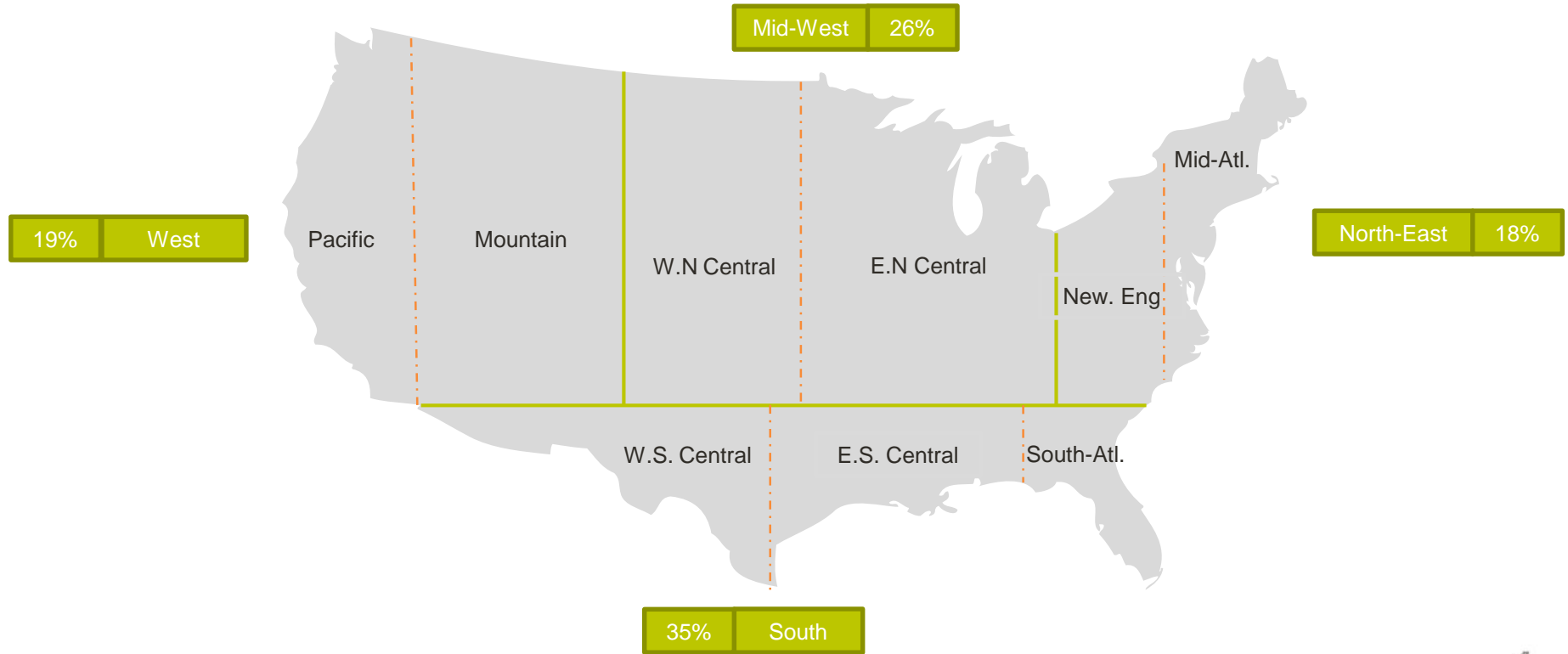


**Third Step**

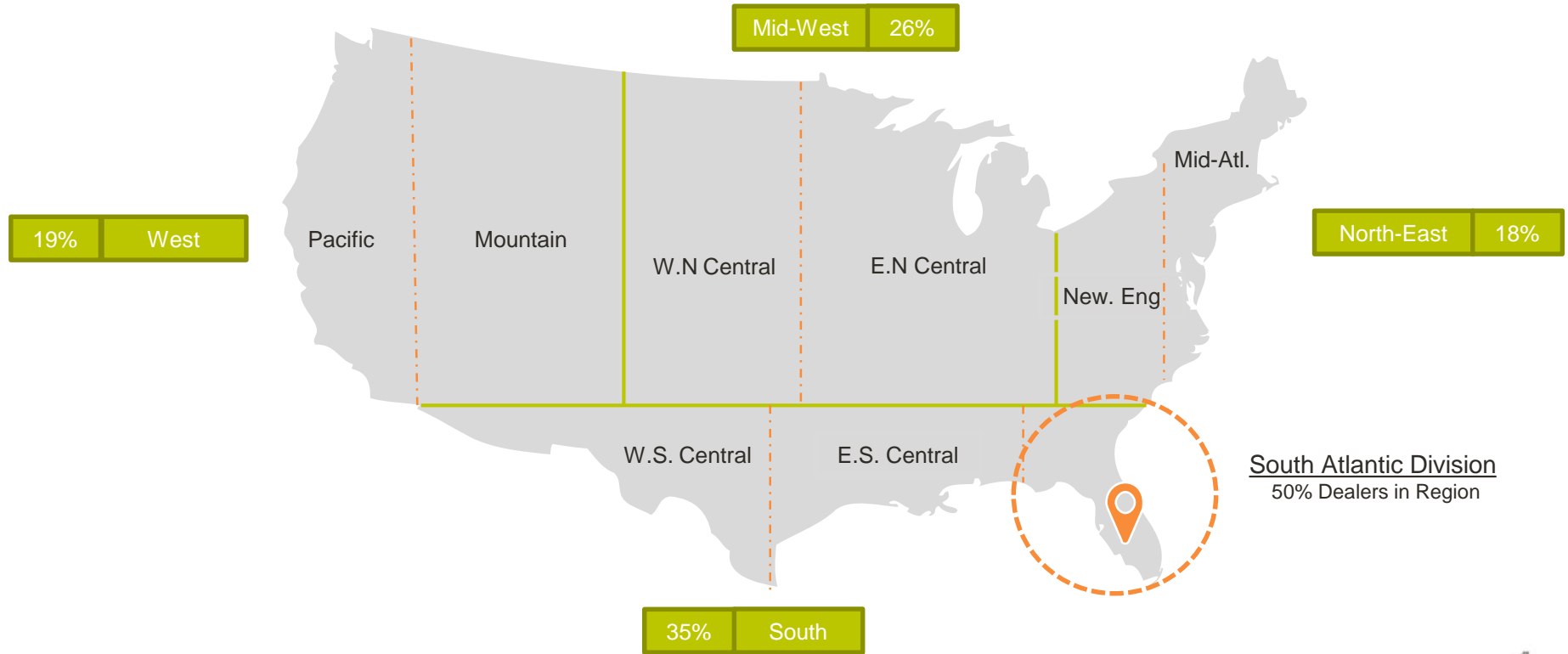


Choosing representative sample for evaluation

# Most Valuable Dealers



# Most Valuable Dealers

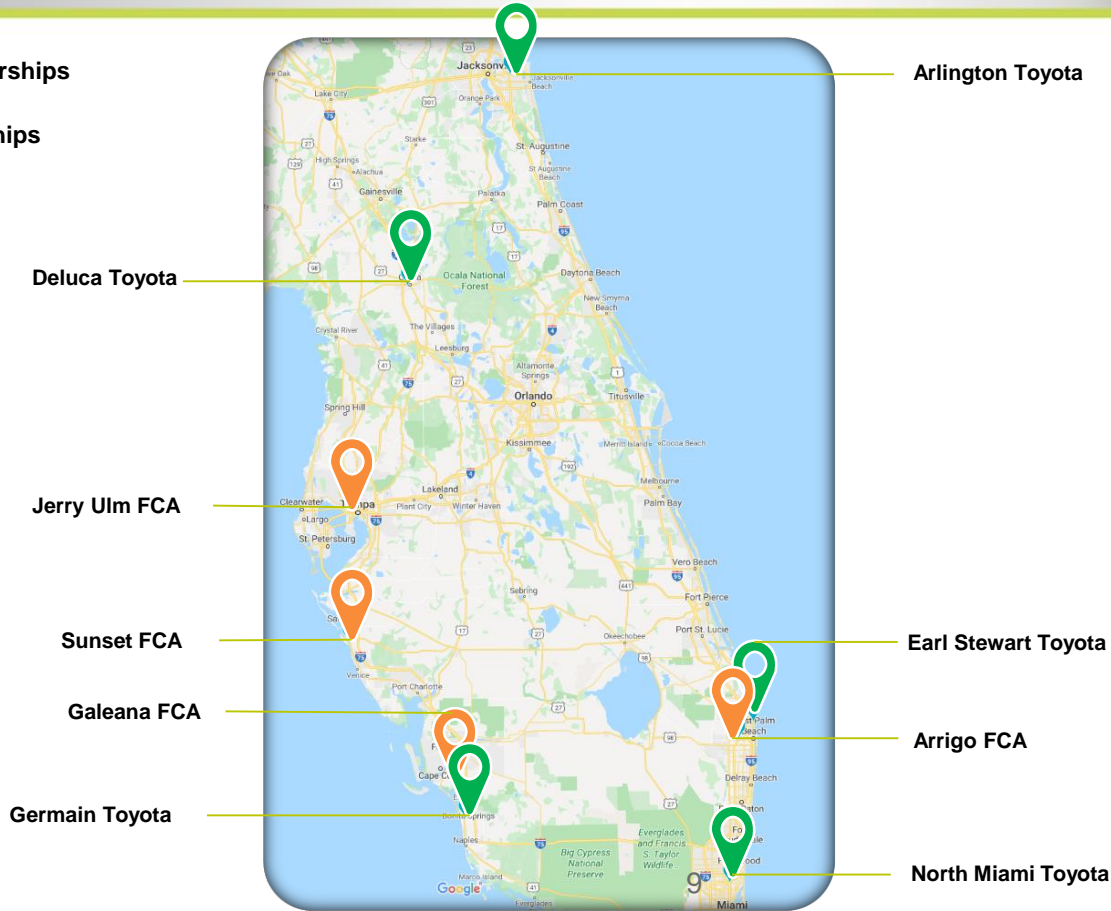




# Most Valuable Dealers

■ Toyota Dealerships

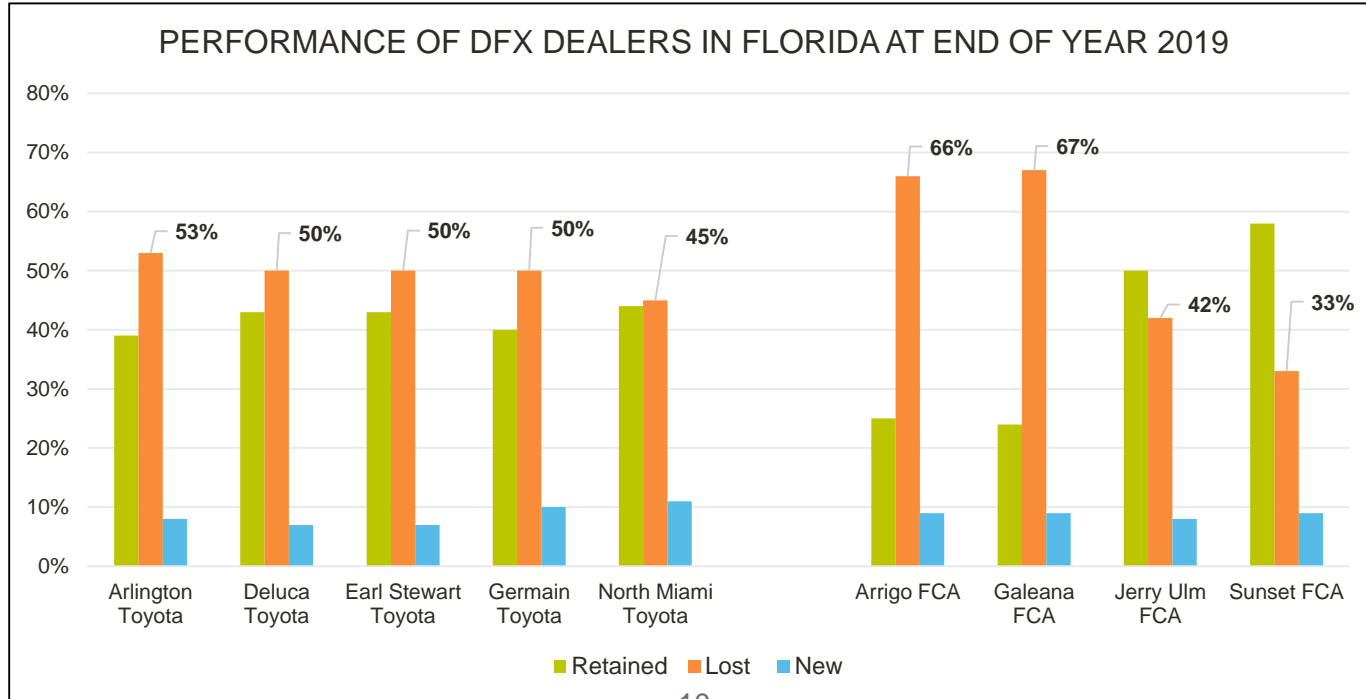
■ FCA Dealerships



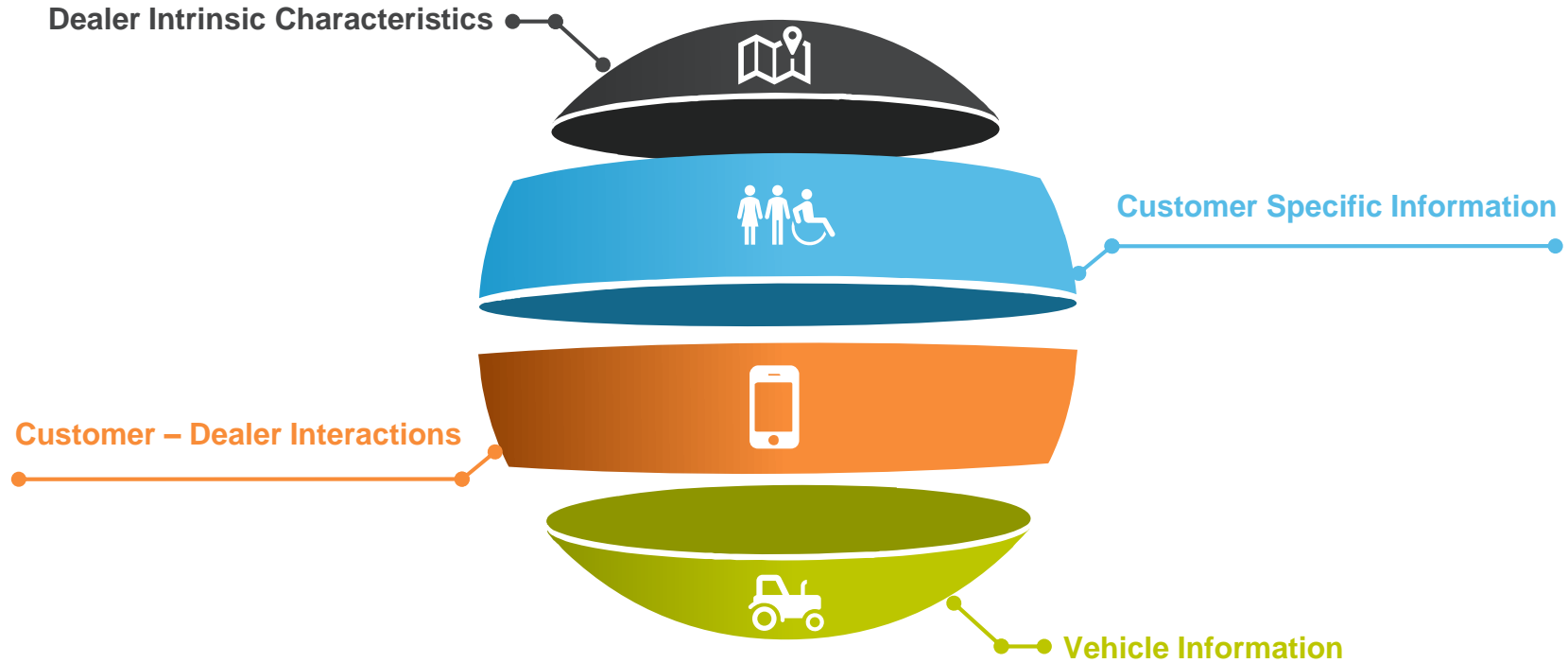
➤ Most Valuable Dealers selected based on **Annual RO Volume**

# Recent Dealer Performance

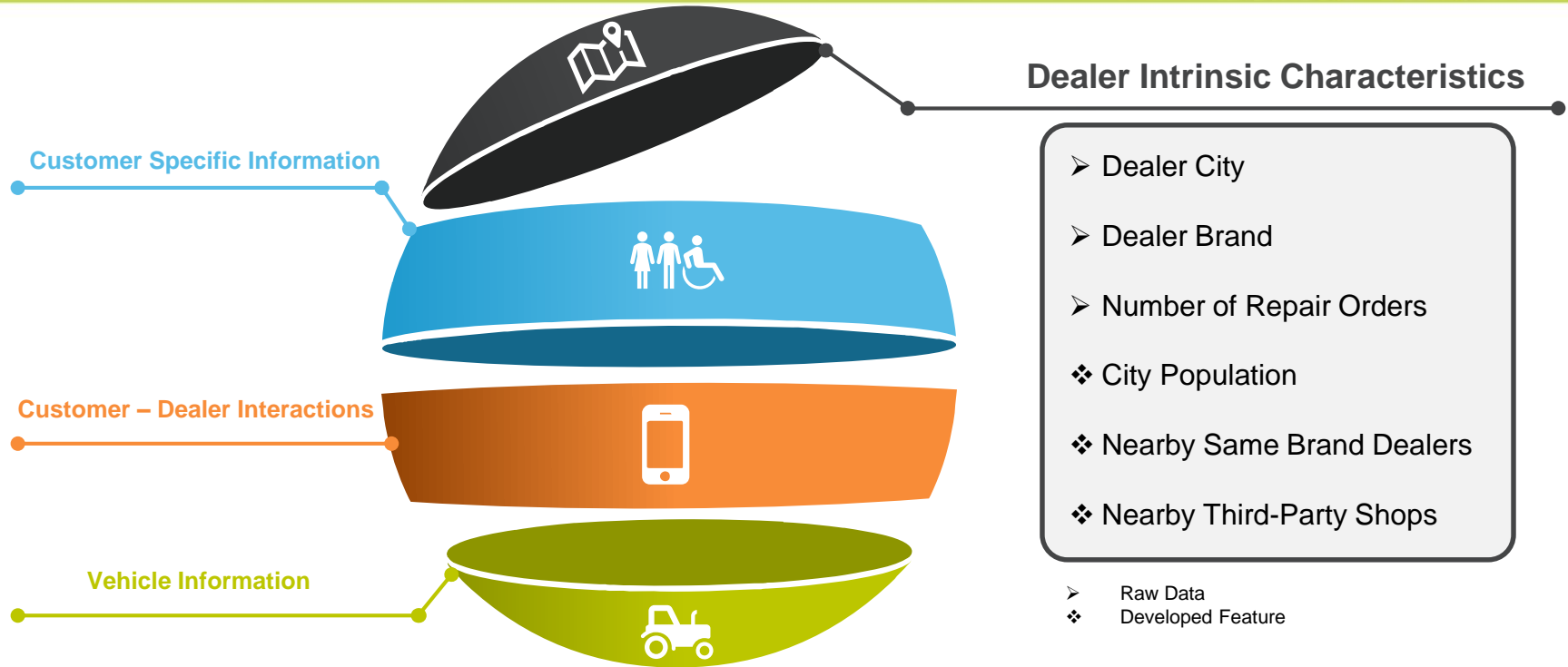
- Toyota Dealers typically outperform FCA Dealers in customer retention by **3%**
- All selected dealers on average have **10%** new customers annually



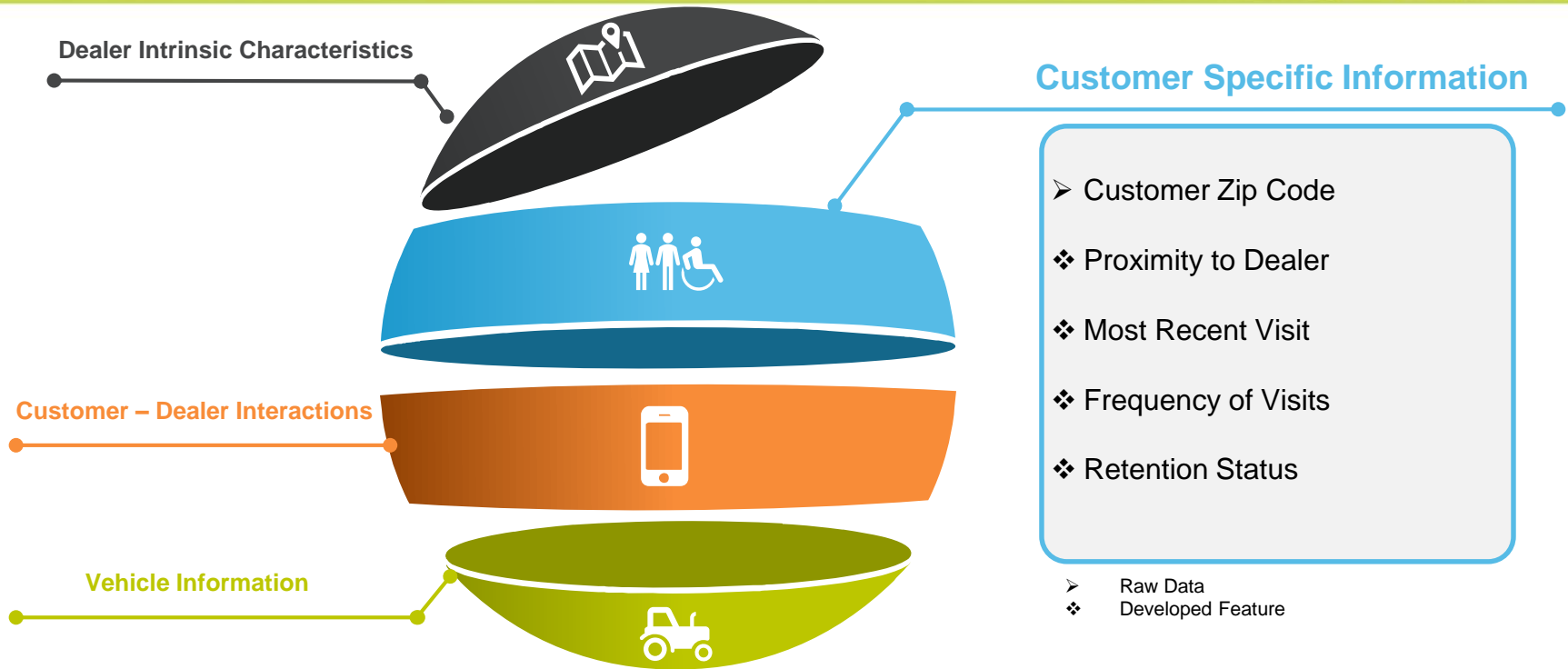
# What is Under the Hood?



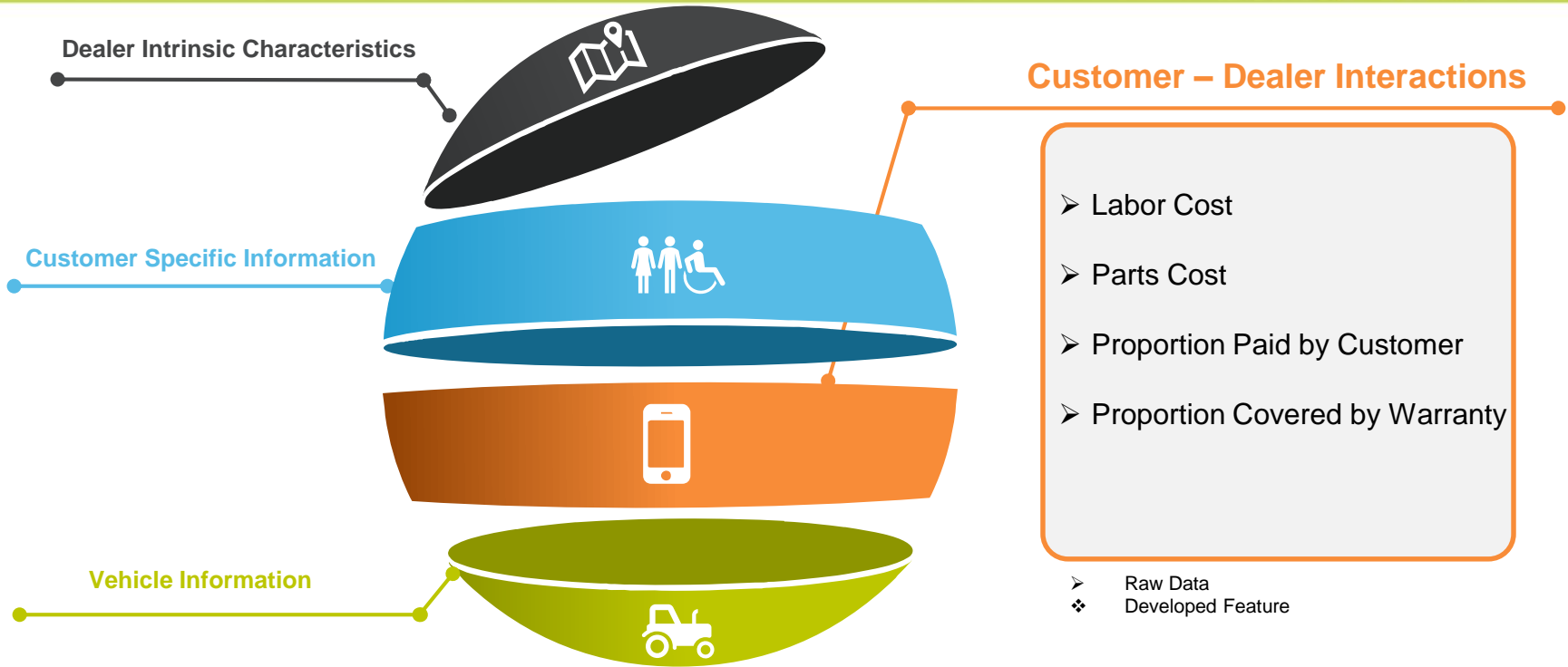
# Defining the Dealership



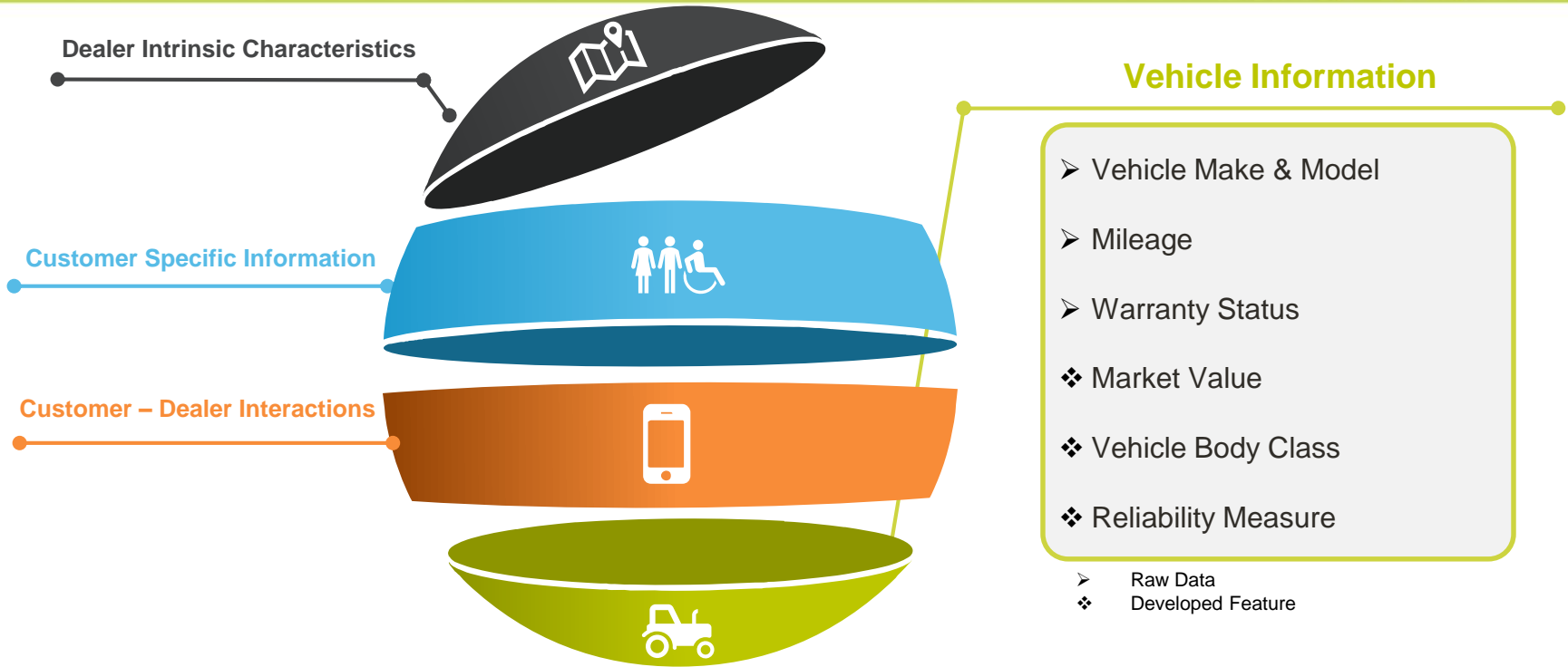
# Understanding the Customer



# Quantifying the Relationship



# Explaining the Vehicle



# Customer Churn Rates



First Step



Predictive Model Development



Second Step



Model Performance Evaluation



Third Step



Understanding Variable Influence



# Predictive Model

## Benefits

- Ensemble models build upon one another to reduce overall error rate
- Minimal need for extensive hyper-parameter tuning

## Method

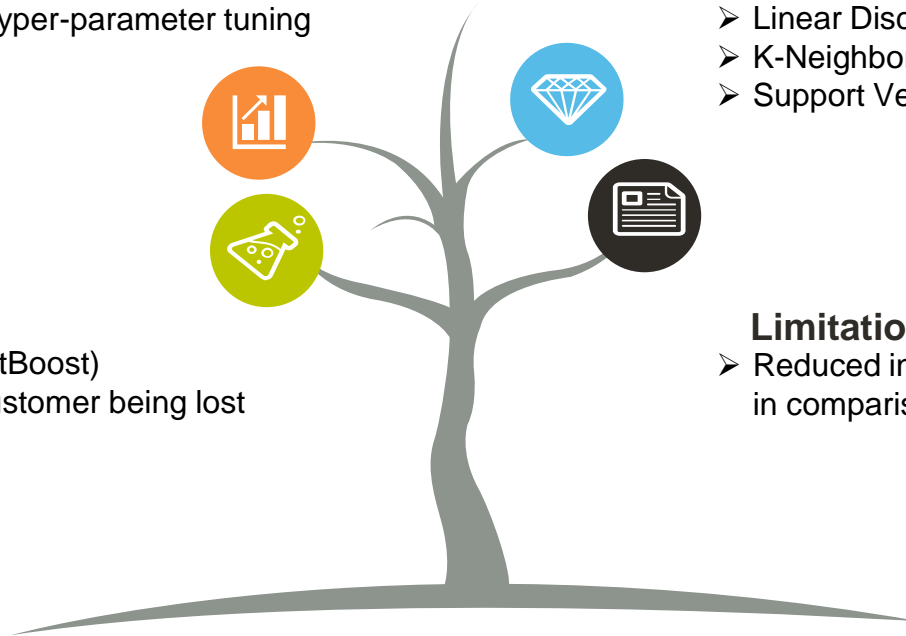
- Ensemble Decision Tree (CatBoost)
- Predict the probability of a customer being lost

## Model Selection

- Logistic Regression
- Ridge Regression
- Linear Discriminant Analysis
- K-Neighbors Classifier
- Support Vector Machine

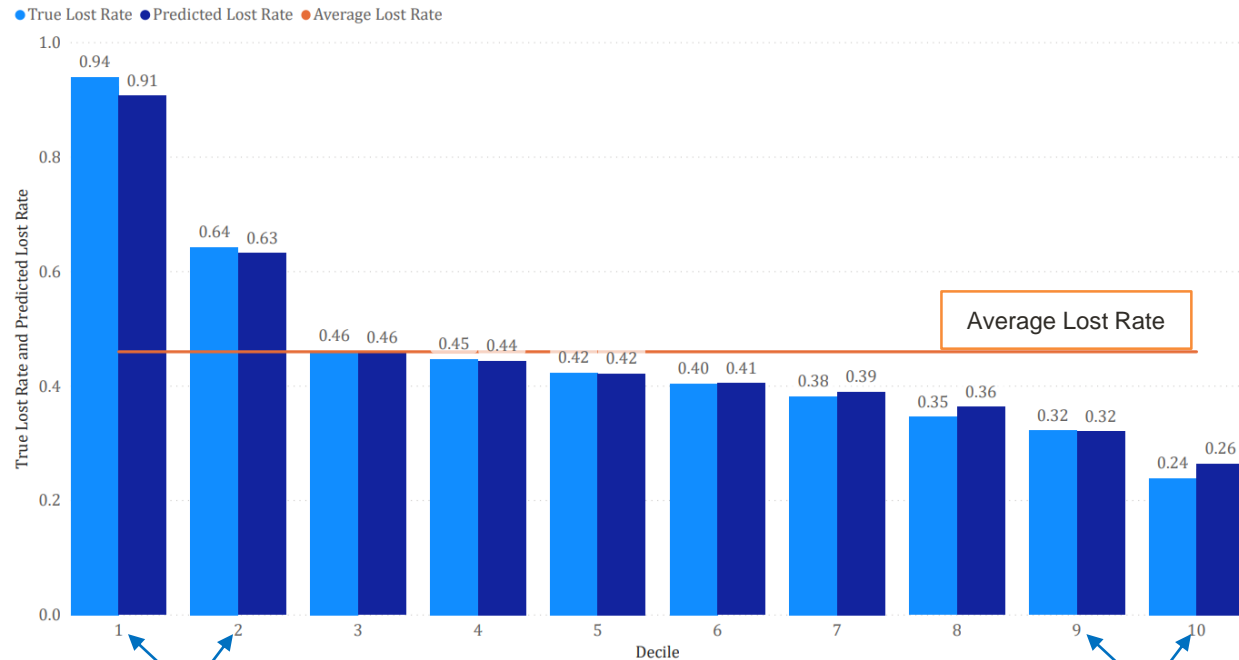
## Limitations

- Reduced interpretability of raw model output, in comparison to alternatives



# Model Performance

Waterfall Plot



Top 20% Lost Risk

Low 20% Lost Risk

➤ Random Customer Lost Probability **46%**

Decile 1 – Top Riskiest Customers:

- Model Prediction of Lost Probability **91%**
- True Lost Rate **94%**
- Outperforms random assignment by **2x**

Decile 2 – Second Riskiest Customers:

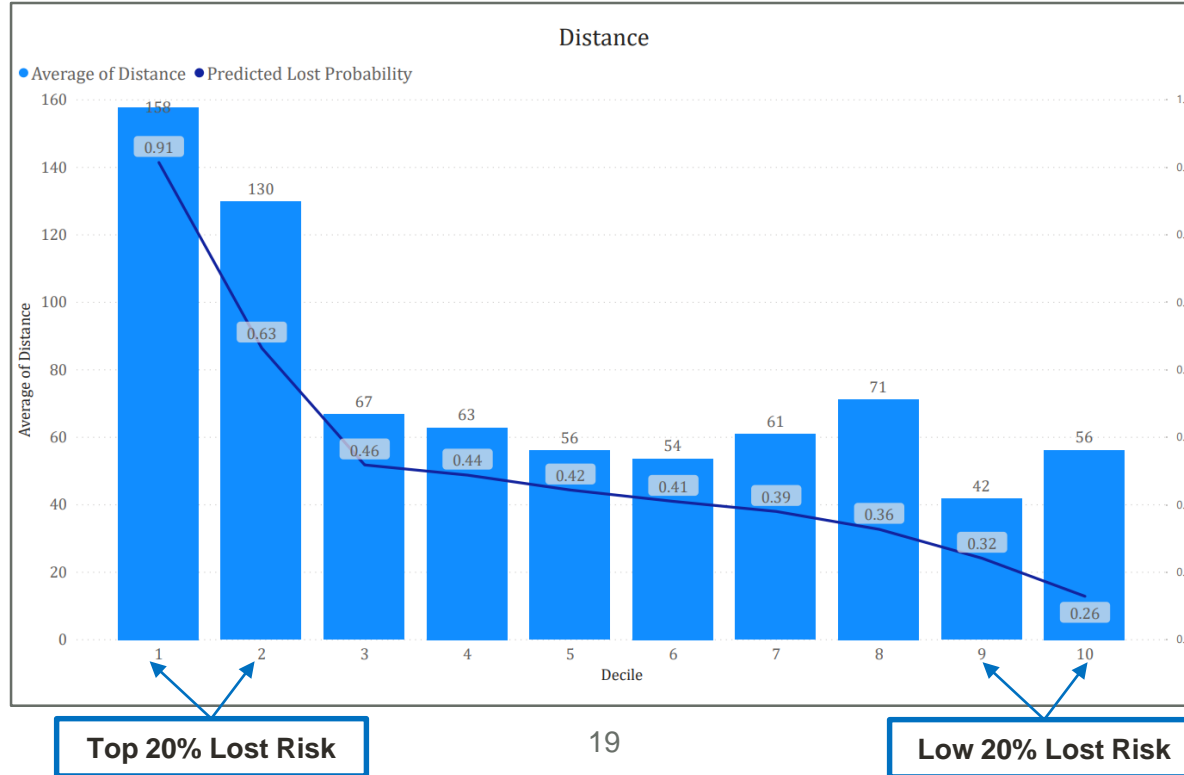
- Model Prediction of Lost Probability **63%**
- True Lost Rate **64%**
- Outperforms random assignment by **1.5x**

➤ Recommend to target **Top 20% Lost Risk Customers**

# Top Features Predicting Risk

## Distance between Customer & Dealership

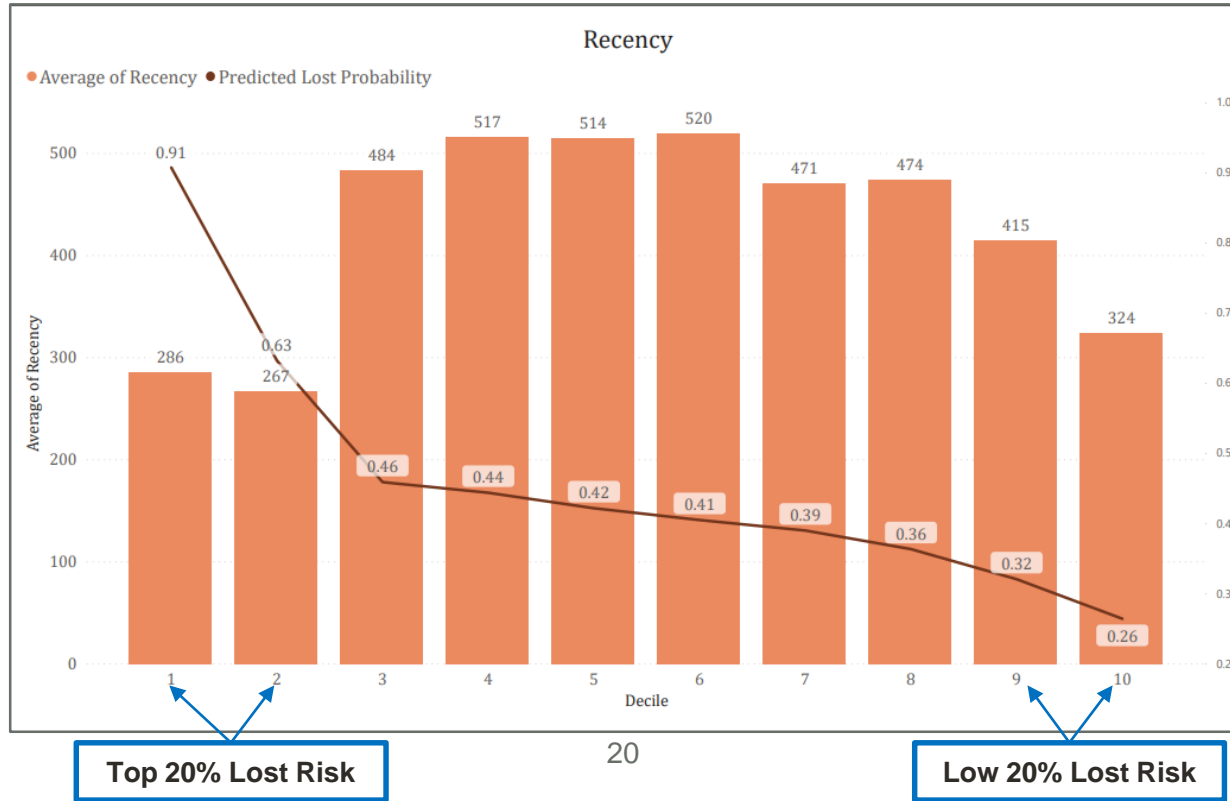
The further away the customer is from the dealership, the higher the risk of losing that customer



# Top Features Predicting Risk

Recent Customer Visit

Riskiest customers have different behaviors than the remaining customers



# Top Features Predicting Risk

## Proportion of Service Paid by Customer

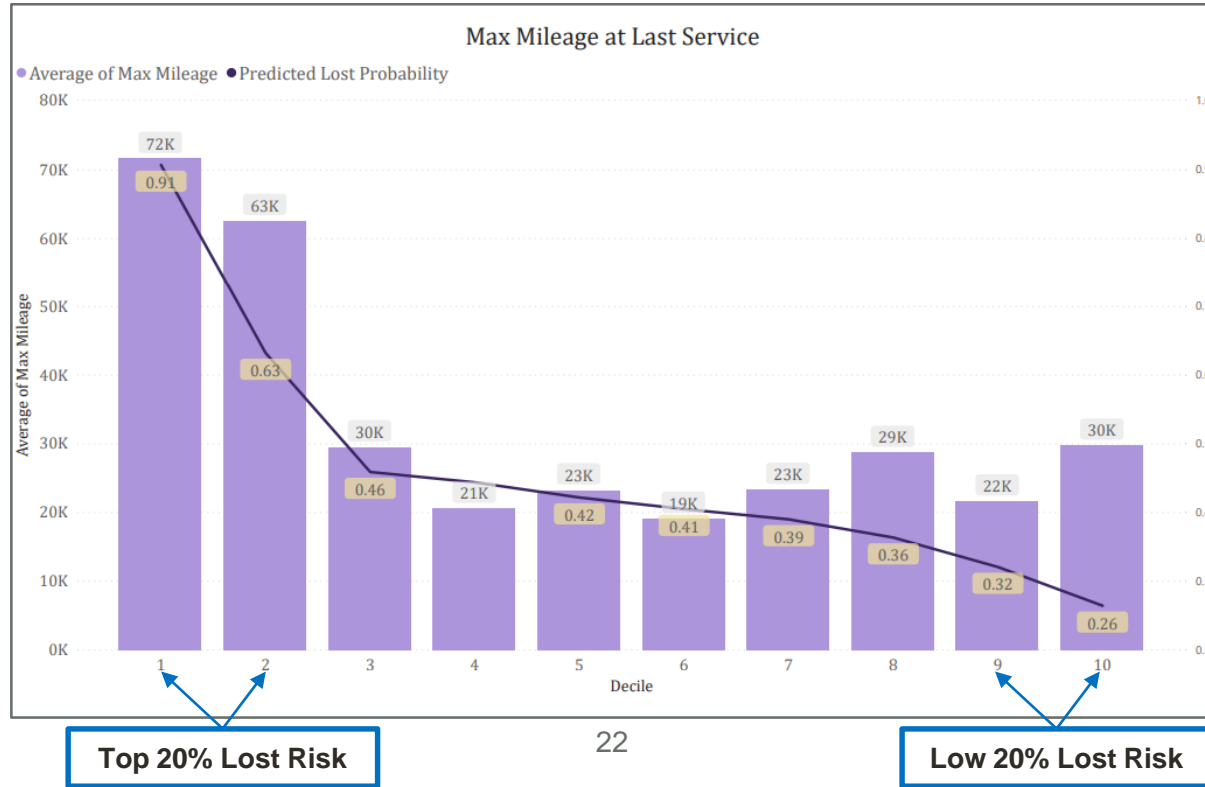
The larger the proportion of *'out of pocket pay'* in a service, the higher the risk of losing that customer



# Top Features Predicting Risk

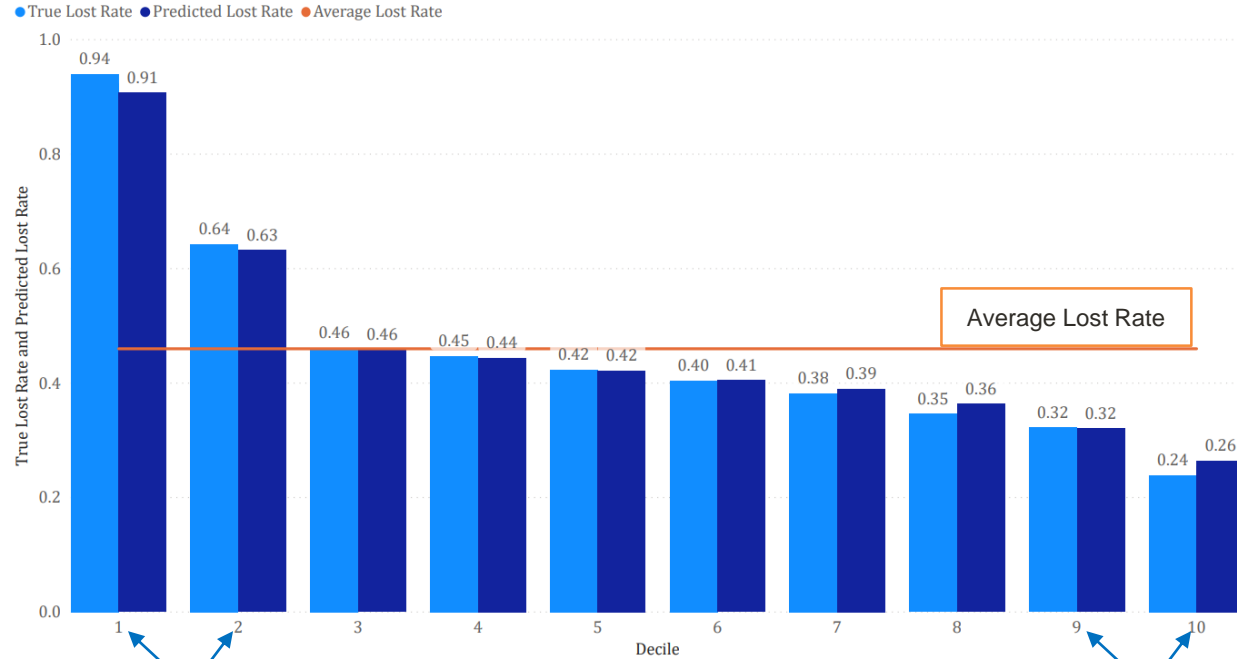
## Mileage at Last Service

Owners of vehicles with higher mileage at last service, are considered a higher lost risk



# Target Recommendation

Waterfall Plot



➤ Recommend to target **Top 20% Lost Risk Customers**

Top 20% Lost Risk

Low 20% Lost Risk

# Attrition Reasons



**First Step**



Difference between Top & Bottom 20% Risk Customers



**Second Step**



The Overall Customer Story



**Third Step**

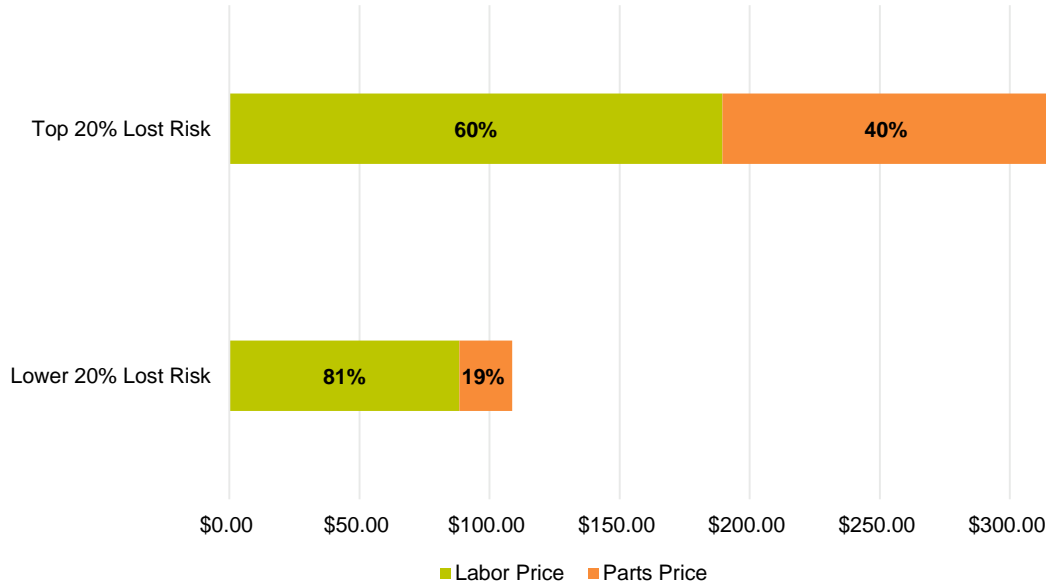


Dealerships in Jeopardy



# How much does service cost?

Average Vehicle Service Costs vs. Customer Risk Deciles



- Top 20% risk customers on average have service bills that are **2.9 times** more expensive than the lower 20%
- Risky customers spend on average **2 times** more the amount for labor & **6 times** more the amount on parts required for service
- Lower 20% risk customers have **30%** more of their service bill dedicated towards labor

# The Customer Story

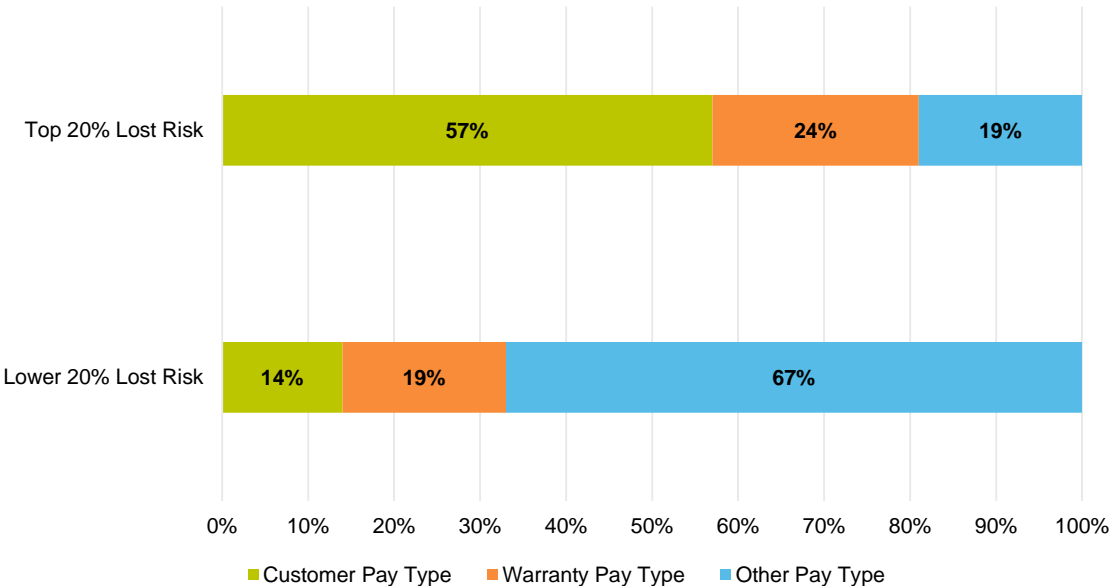


## Overall Service Costs

Riskier customers typically bring their vehicles for major service & parts replacements, as oppose to general scheduled maintenance

# How much do customers spend?

Service Payment Type vs. Customer Risk Deciles



- Top 20% risky customers are on average **4 times** more likely to be paying out of their own pocket
- Riskier customers have around **30%** more warranty claims

\*\*Other Pay Type includes entries not related to either Customer/Warranty Pay

# The Customer Story



1

## Service Payment Type

Riskier customers are paying for their own service expenses, and generally have more frequent warranty claims

## Overall Service Costs

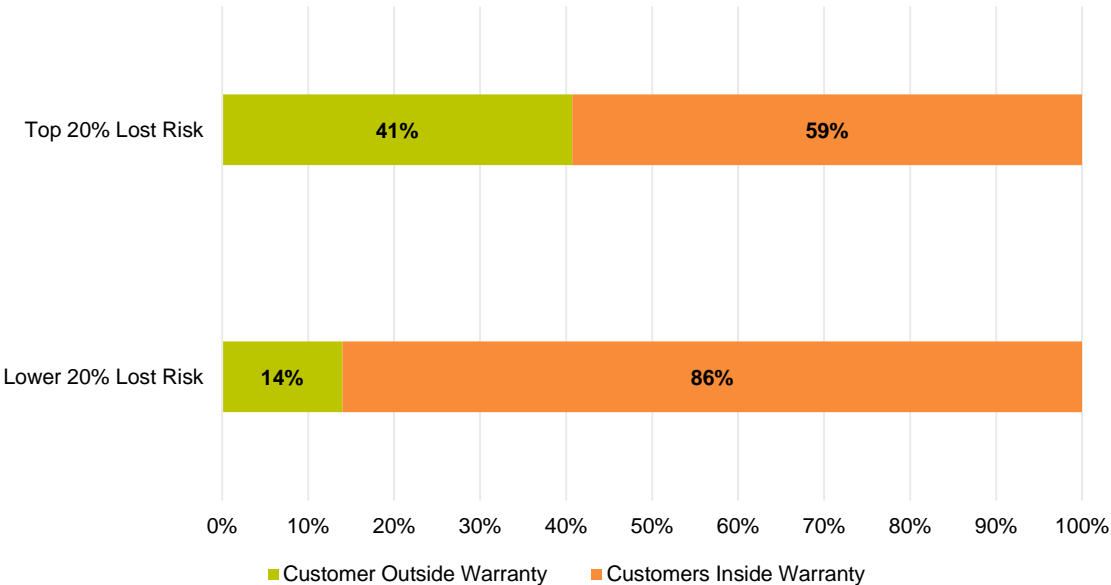
Riskier customers typically bring their vehicles for major service & parts replacements, as oppose to general scheduled maintenance

2



# What is the vehicle condition?

Customer Warranty Status vs. Customer Risk Deciles



- Top 20% risky customer are **3 times** more likely to be outside of warranty
- Riskier customers own vehicles that are **3 years** older and have a market value that is around **\$8,000** cheaper

	Average Warranty Mileage	Average RO Vehicle Mileage	Average Vehicle Age	Average Market Value
Top 20% Loss Risk	71,000	67,000	5.9	\$ 14,444.00
Lower 20% Loss Risk	76,000	24,000	2.9	\$ 22,394.00

# The Customer Story



1

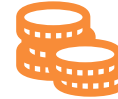
## Overall Service Costs

Riskier customers typically bring their vehicles for major service & parts replacements, as oppose to general scheduled maintenance

## Service Payment Type

Riskier customers are paying for their own service expenses, and generally have more frequent warranty claims

2



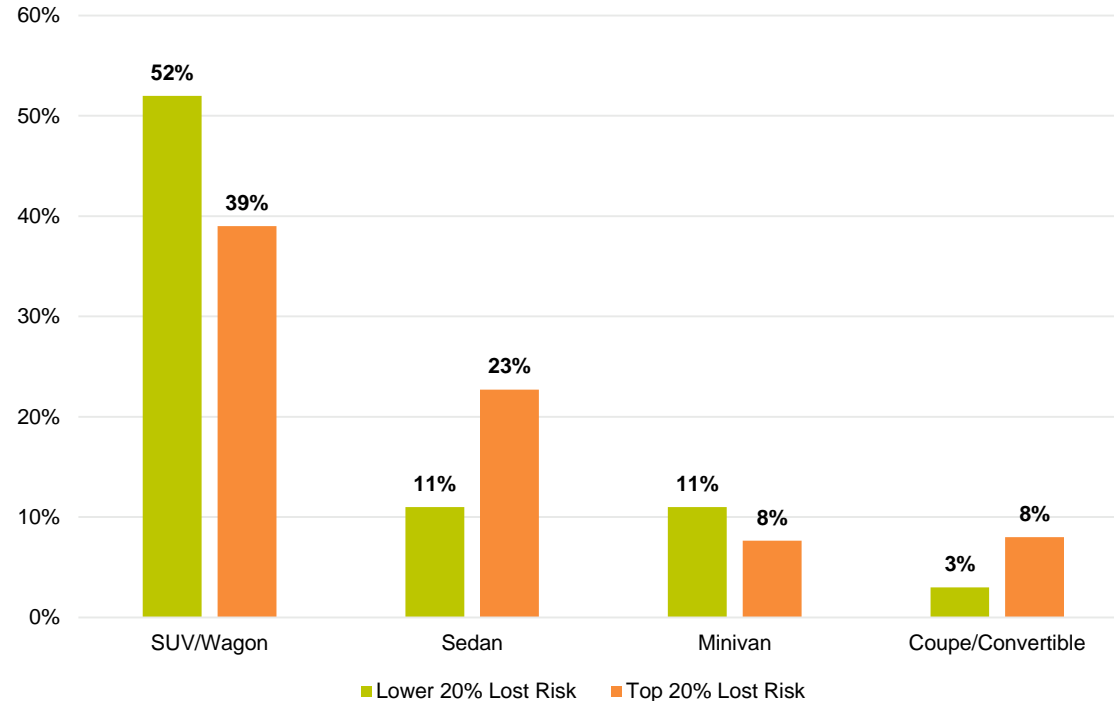
3

## Vehicle Condition

Riskier customers tend to have older & cheaper vehicles that are nearing the end of their warranty period

# What car do they drive?

Vehicle Body Class vs. Customer Risk Deciles



- Top 20% risky customer are **2 times** more likely to own a Sedan vehicle
- Convertible/Coupe owners are **2.5 times** more likely to be risky customers
- Around **37% more** Minivan owners are considered less risky
- SUV/Wagon vehicles compromise the largest market share of **40%**

# The Customer Story



1

## Overall Service Costs

Riskier customers typically bring their vehicles for major service & parts replacements, as oppose to general scheduled maintenance

## Service Payment Type

Riskier customers are paying for their own service expenses, and generally have more frequent warranty claims

2



3

## Vehicle Condition

Riskier customers tend to have older & cheaper vehicles that are nearing the end of their warranty period

## Vehicle Type

Riskier customers will typically drive: *Sedans, Convertibles/Coupes*; whereas less risky customers will tend to drive *SUV/Wagons*

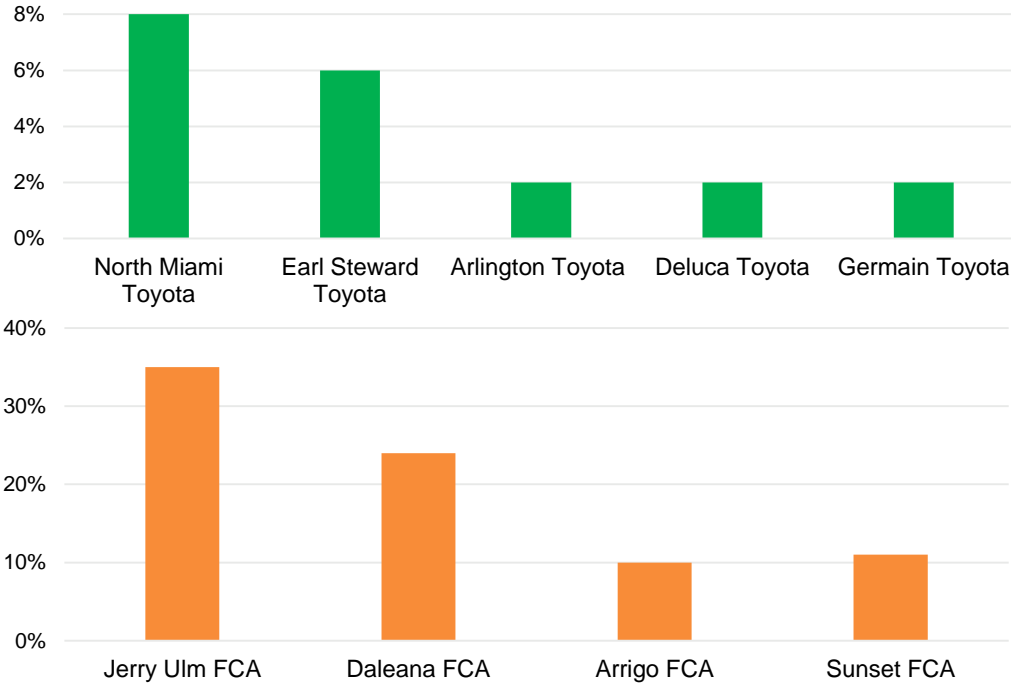
4





# Dealerships in Jeopardy

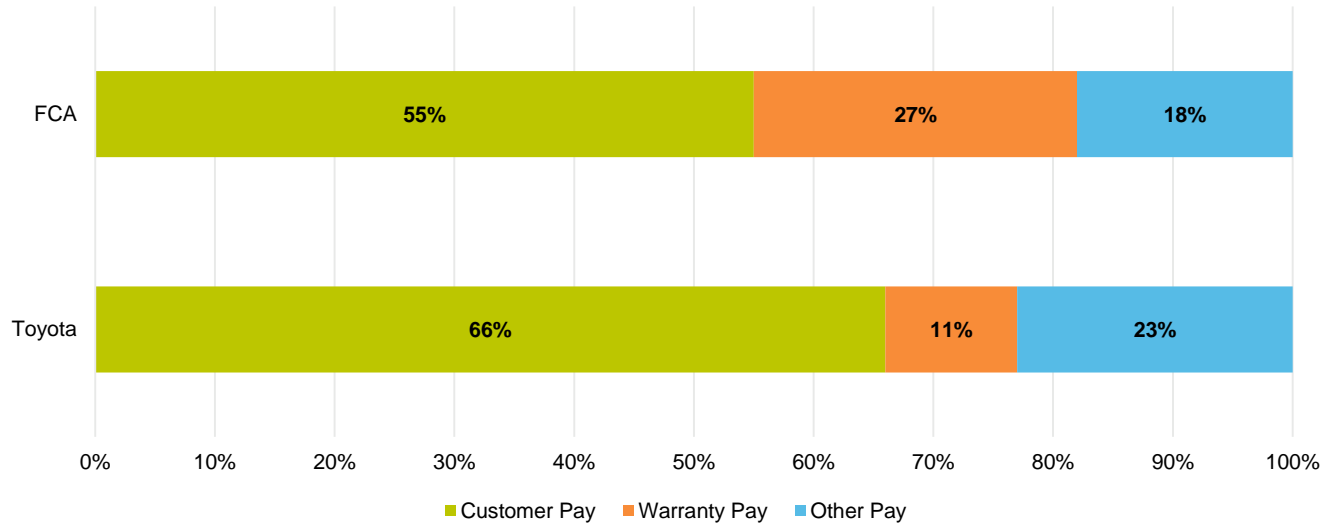
Top 20% Risk Customers in DFX Florida Dealers



- Toyota dealers have approximately **twice** the number of repair orders
- FCA dealers in Florida are on average **5 times** riskier than Toyota dealers
- Toyota dealers have vehicles with an average reliability score of **72**, and FCA **58** (range 44 – 81)

# Dealerships in Jeopardy

Service Payment Type of Top 20% Risky Customers per Dealership Brand



- Riskier customers visiting FCA Dealers are **twice** more likely to perform warranty service
- Risky Toyota Customers are **1.2 times** more likely to pay for their service

\*\*Other Pay Type includes entries not related to either Customer/Warranty Pay

# Implementation



**First Step**



Customer Score Card Deliverable



**Second Step**



Contact List Deliverable



**Third Step**



Estimated Cost-Benefit of a Marketing Campaign

# Customer Score Card

## Very High Risk

Convertible/Coupe Vehicles

Mkt. Value - \$24,000

Avg. Age – 3.5 Years

Avg. Spend - \$190

Customer Spend – 40%

Warranty Coverage – 17%

## Medium Risk

Sedan Vehicles

Mkt. Value - \$17,500

Avg. Age – 4.3 Years

Avg. Spend - \$218

Customer Spend – 33%

Warranty Coverage – 11%

## Largest Mkt. Share

SUV/Wagon Vehicles

Mkt. Value - \$22,900

Avg. Age – 2.9 Years

Avg. Spend - \$175

Customer Spend – 23%

Warranty Coverage – 20%

- Pre-developed Segmentation Criteria based on choice of characteristics
- Customers are *Tagged* based on their proximity to a certain score card
- Allows for continuous evaluation of top risk customers for pro-active targeting

# Contact List

DFX Analytics Team



Customers in Top  
20% Risk Group



Jerry Ulm FCA

DFX Marketing Team



Customer List to Contact at Jerry Ulm FCA

Customer ID	VIN	Vehicle Info	Recent Visit (Days)	Avg. Service Price (\$)	Lost Probability
72288385	1C6RR7NT8GS329978	2016 RAM 1500 (5.7L)	721	\$129.75	99.8%
83519155	1C6RD6LT9CS333415	2012 RAM 1500 (5.7 L)	345	\$165.20	99.7%
74115321	1C6RR6KG4HS683287	2016 Jeep Cherokee	380	\$253.35	87.4%
70616663	2C3CDXBG7HH525377	2017 Dodge Charger (3.6L)	689	\$237.86	86.83%

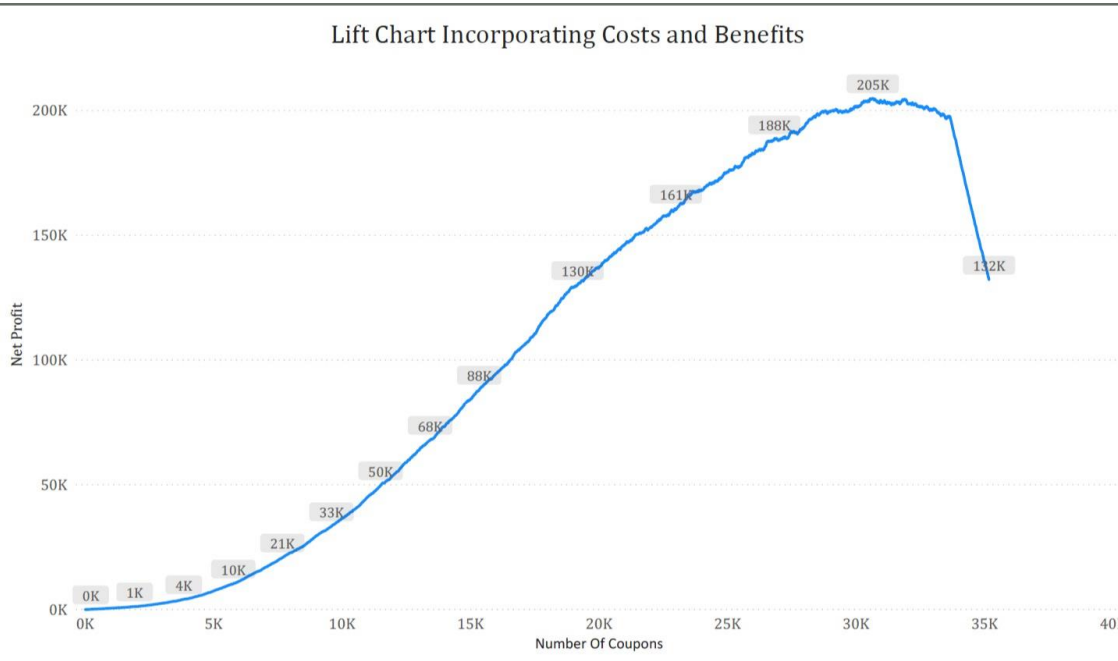
\*\* Sample Results based on Dealer ID 5961, Jerry ULM FCA

DFX Dealerships



# Targeting Costs

Lift Chart Incorporating Costs and Benefits



\*\* Mock up Results based on Assumptions

## Assumptions:

- Revenue generated from retaining one of the Top 20% Risk Customers is **\$190**
- Discount provided by the Dealership throughout the campaign is **35%**

## Outcome:

- The Cut-off probability for sending out discounts is **56%**
- The Optimal number of customers to target per marketing campaign is around **30,000**

# Future Work

- Capture vehicle ownership changes
  - Analyze Voluntary vs. Involuntary Churners
- Expand Churn Analysis to Regional Level
  - Use same features
  - Implement model for each subset
- Develop Marketing Campaigns to Top 20% Risk Customers
  - Monitor customer response rate



**Thank You!**

---



# Table of Contents

## Introduction

- Objectives
- Project Setting

## Dealer Selection

- Dealer Segmentation
- Dealer Selection
- Dealer Performance

## Model Features

- Descriptive Characteristics

## Model Analysis

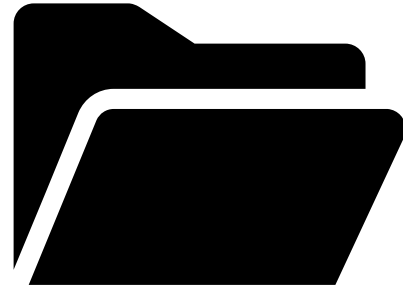
- Predictive Model
- Model Evaluation
- Feature Importance

## Attrition Reasons

- Service Cost
- Service Payment Type
- Vehicle Condition
- Types of Vehicles
- Dealership Risk

## Conclusion

- Score Card Deliverable
- Customer List Deliverable
- Cost-Benefit Analysis
- Future Work



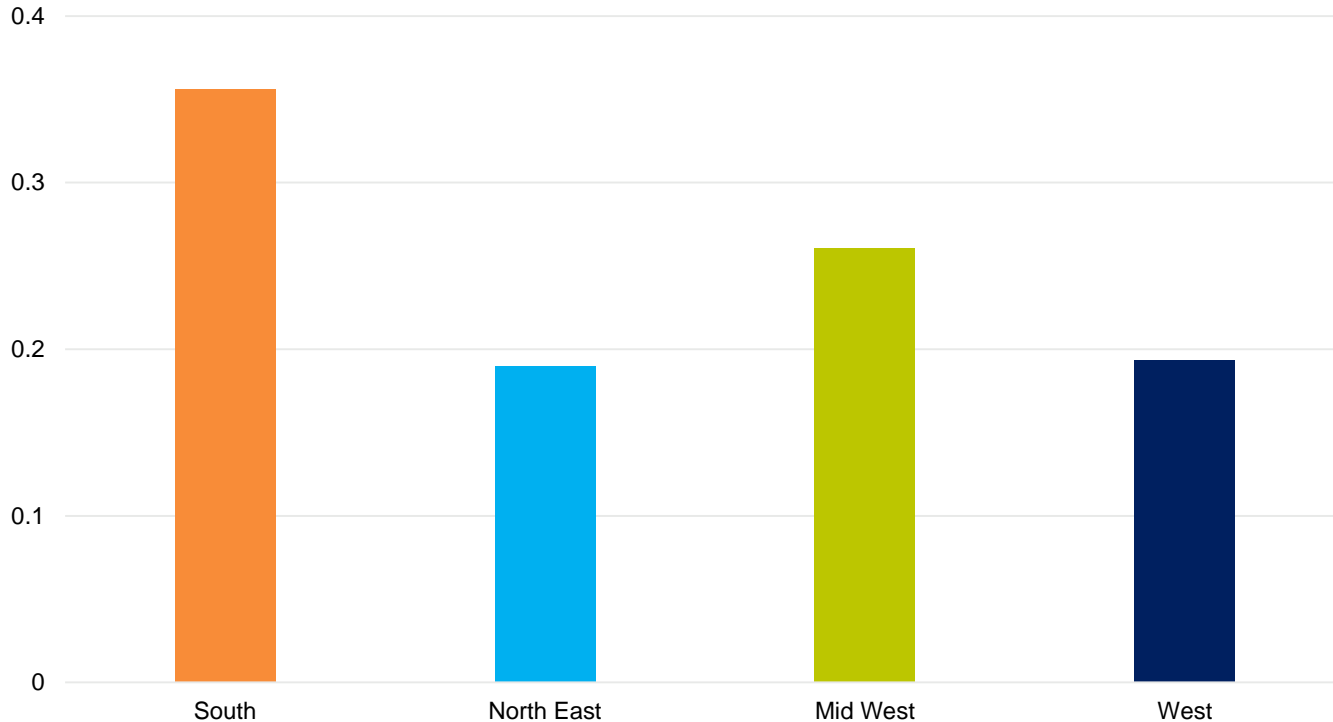


# Appendix

---

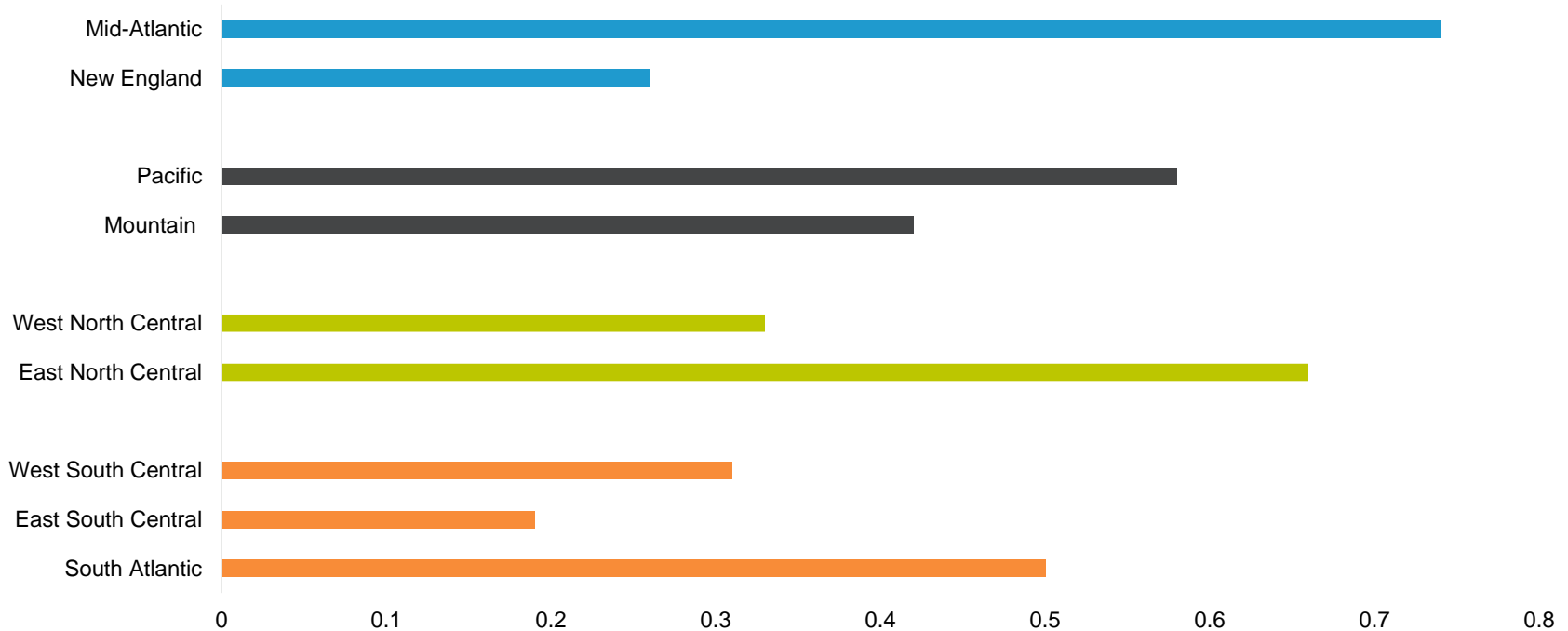
# Appendix – USA Regions

Distribution of Dealers per Division in the USA



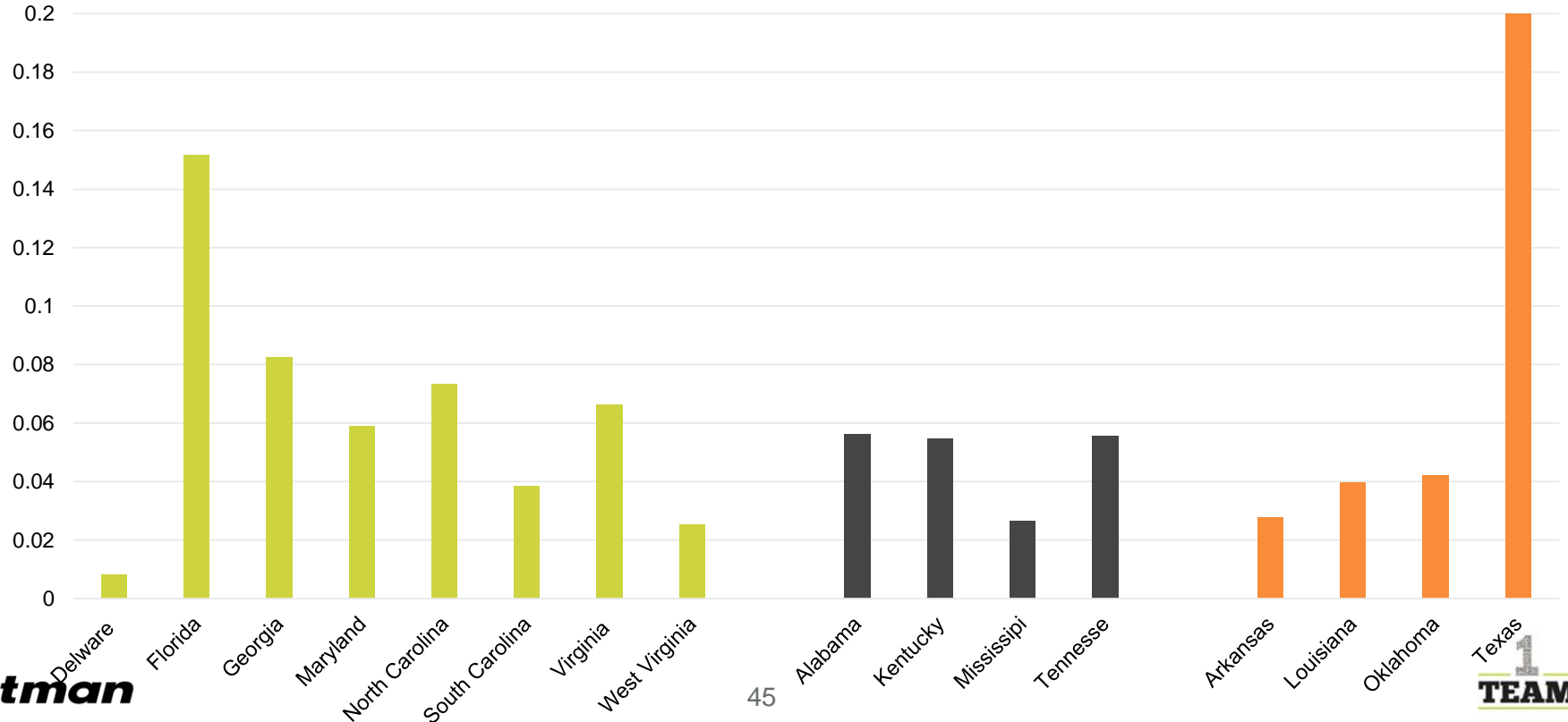
# Appendix – USA Divisions

Distribution of Dealers per Division in the USA



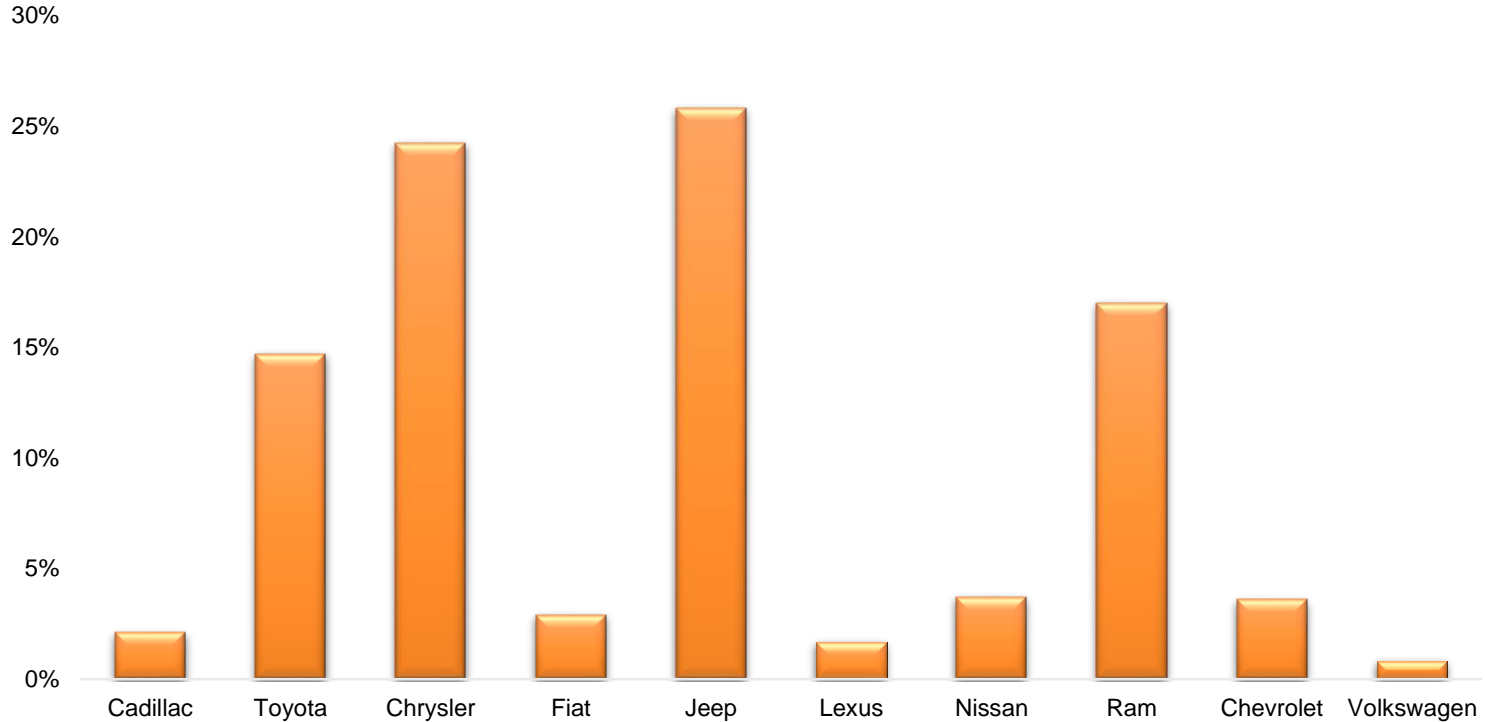
# Appendix – South Region

Distribution of Dealers in South Region



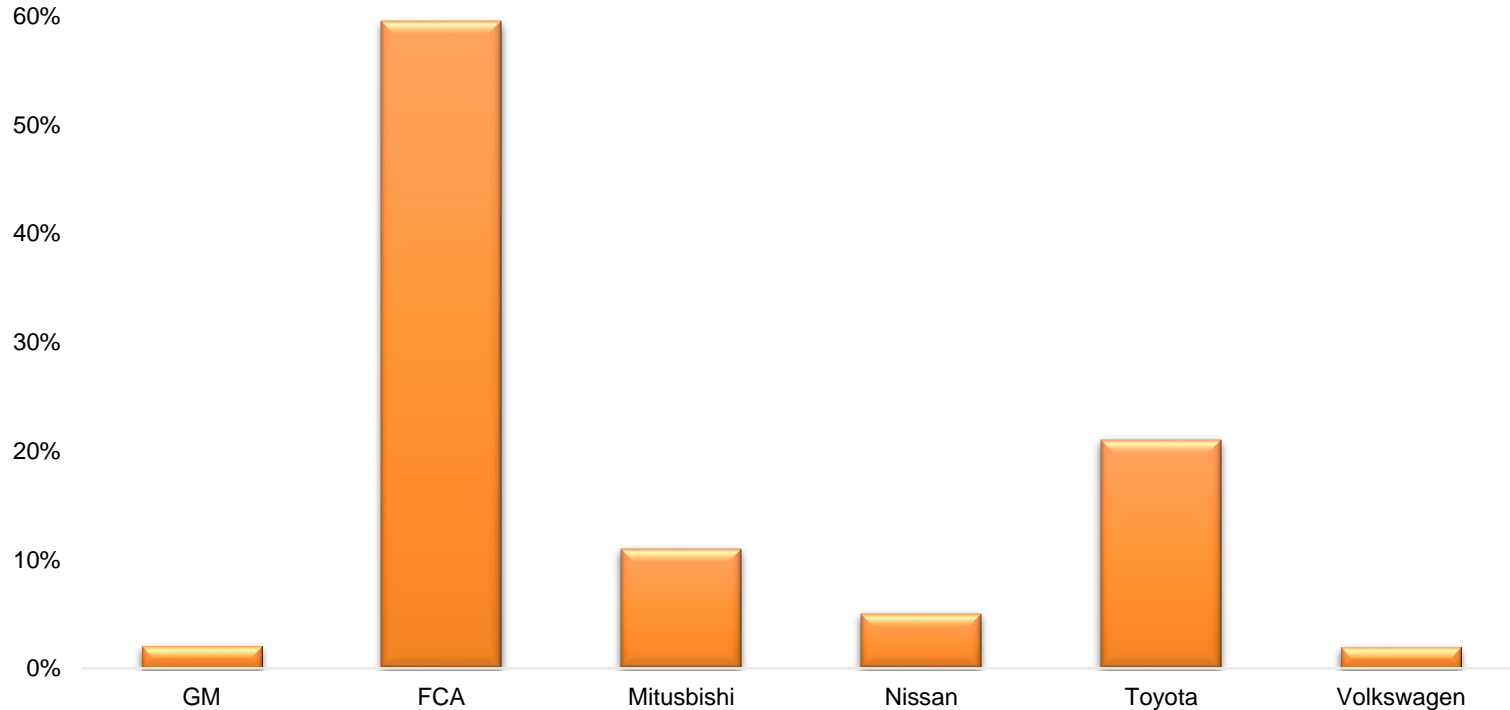
# Appendix - Dealer Grouping

Distribution of DFX Vehicle Makes in Florida



# Appendix - Dealer Grouping

Distribution of DFX Dealer Groups in Florida



# Appendix – Testing Results

MMA Class of 2020

12/12

100% of 8 values



	minProb	maxProb	numOfCustomer	cumNumOfCustomer	trueLoss	trueLossRate	cumTrueLoss	cumTrueLossRate	predLossRate	actuaLlft	predictLift	decile
0	0.780503	0.998242	5277	5277	4774	0.904681	4774	0.904681	0.907007	1.963130	1.968179	1
1	0.485389	0.780423	5275	10552	3246	0.615355	8020	1.520036	0.633539	1.335303	1.374762	2
2	0.450288	0.485388	5276	15828	2322	0.440106	10342	1.960142	0.459768	0.955017	0.997684	3
3	0.434944	0.450287	5278	21106	2390	0.452823	12732	2.412965	0.444226	0.982613	0.963956	4
4	0.412468	0.434938	5276	26382	2172	0.411676	14904	2.824641	0.422438	0.893324	0.916679	5
5	0.398594	0.412465	5277	31659	2114	0.400606	17018	3.225247	0.405416	0.869304	0.879741	6
6	0.381073	0.398590	5276	36935	2089	0.395944	19107	3.621191	0.390614	0.859187	0.847622	7
7	0.335387	0.381070	5277	42212	1946	0.368770	21053	3.989961	0.364716	0.800220	0.791423	8
8	0.309818	0.335378	5276	47488	1676	0.317665	22729	4.307626	0.321331	0.689323	0.697279	9
9	0.015074	0.309808	5277	52765	1587	0.300739	24316	4.608365	0.266506	0.652595	0.578310	10

Team 9 - Winter Term

Monday

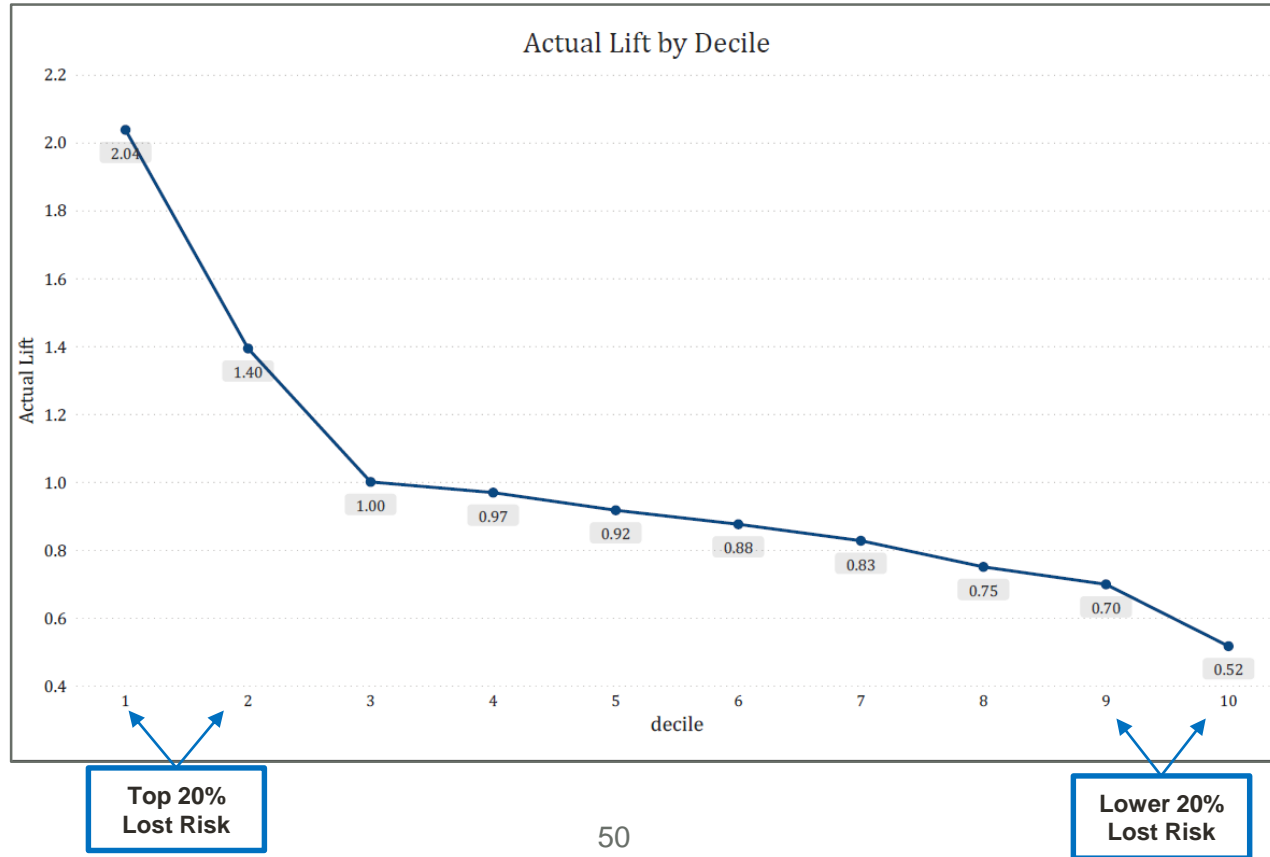
12/12



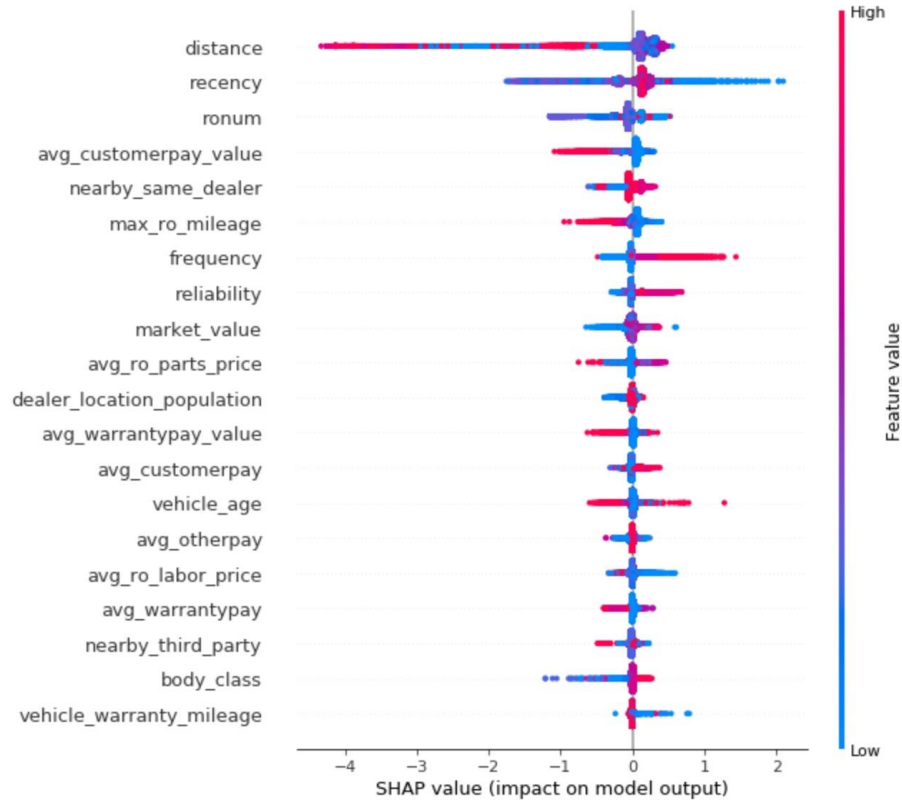
# Appendix – Alternative Models

0	Light Gradient Boosting Machine	0.6445
1	CatBoost Classifier	0.6444
2	Extreme Gradient Boosting	0.642
3	Gradient Boosting Classifier	0.6416
4	Ada Boost Classifier	0.6264
5	Ridge Classifier	0.6039
6	Linear Discriminant Analysis	0.6034
7	Logistic Regression	0.5853
8	Random Forest Classifier	0.5764
9	Naive Bayes	0.5723
10	Extra Trees Classifier	0.5662
11	K Neighbors Classifier	0.5612
12	Decision Tree Classifier	0.5516
13	Quadratic Discriminant Analysis	0.5377
14	SVM - Linear Kernel	0.4955

# Appendix - Model Evaluation

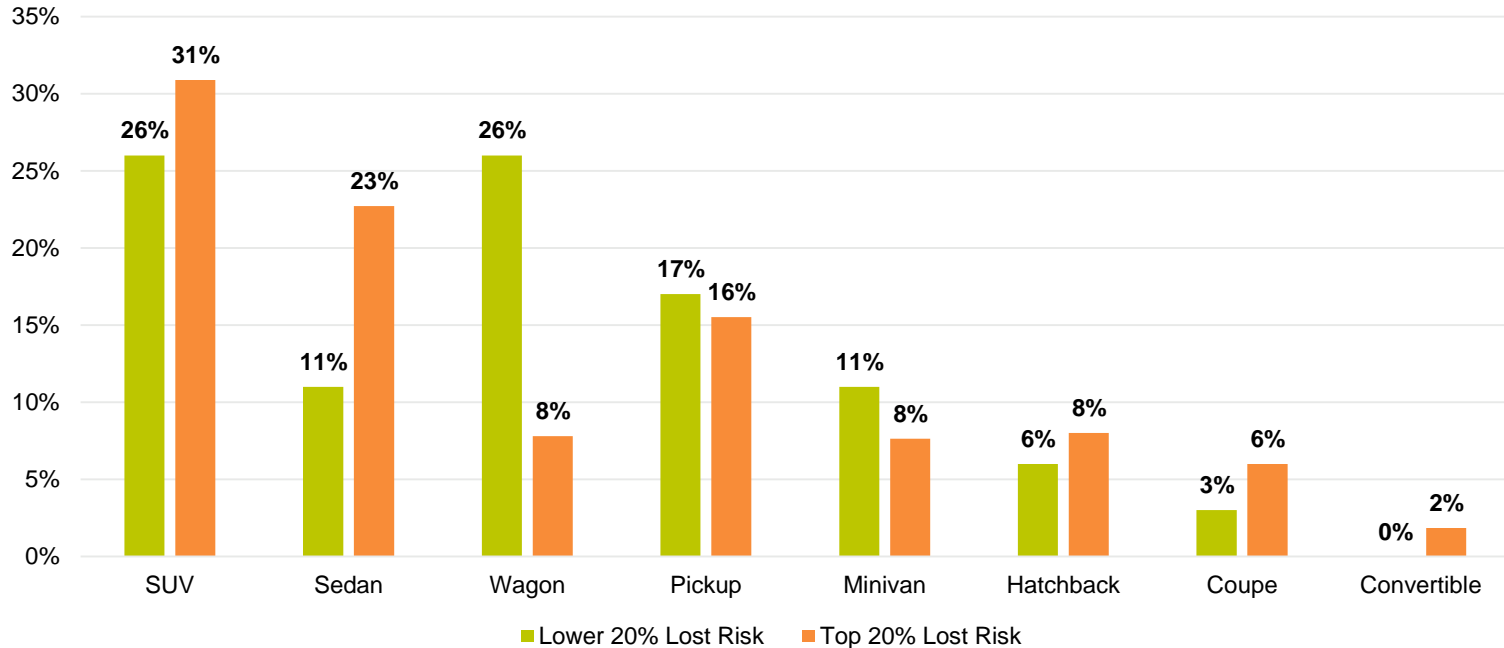


# Appendix - Feature Importance



# Appendix – Vehicle Type

## Vehicle Body Class vs. Customer Risk Deciles



# Appendix – Vehicle Type

## SUV/Wagon Models

- ✓ (Select All)
- ✓ 300 SRT (6.4L)
- ✓ 3500 Diesel (6.7L)
- ✓ 4Runner
- ✓ Aspen (4.7L)
- ✓ Aspen (5.7L)
- ✓ Aspen Hybrid (5.7L)
- ✓ FJ Cruiser
- ✓ Highlander
- ✓ Highlander Hybrid
- ✓ Land Cruiser
- ✓ Matrix (1.8L)
- ✓ Pacifica (3.5L)
- ✓ Pacifica (3.8L)
- ✓ Pacifica (4.0L)
- ✓ Prius v
- ✓ Prius V Hybrid (1.8L)
- ✓ ProMaster City Cargo Van (2.4L)
- ✓ ProMaster City Wagon (2.4L)
- ✓ RAV4
- ✓ Sequoia (4.6L)
- ✓ Sequoia (4.7L)
- ✓ Sequoia (5.7L)
- ✓ Sienna
- ✓ Town & Country (3.3L)
- ✓ Town & Country SWB (3.3L)
- ✓ Venza
- ✓ Voyager (2.4L)
- ✓ Voyager (3.3L)

\*\* Sample Models

- Large family sized SUV's
- Medium sized Cross-over vehicles



# Appendix – Vehicle Type

## Hatchback Models

- ☒ (Select All)
- ☒ Celica
- ☒ Matrix (1.8L)
- ☒ Pacifica (3.5L)
- ☒ Pacifica (3.8L)
- ☒ Pacifica (4.0L)
- ☒ Prius
- ☒ Prius c
- ☒ Prius c Hybrid (1.5L)
- ☒ Prius Hybrid (1.5L)
- ☒ Prius Plug-In
- ☒ Prius Plug-In Hybrid (1.8L)
- ☒ PT Cruiser (2.4L)
- ☒ PT Cruiser Classic (2.4L)
- ☒ Yaris

\*\* Sample Models



# Appendix – Vehicle Type

## Coupe/Convertible Models

- ☒ (Select All)
- ☒ 200 (2.4L)
- ☒ 200 (2.4L) 6 Spd
- ☒ 200 (3.6L)
- ☒ 200 (3.6L) 6 Spd
- ☒ Camry Solara
- ☒ Crossfire (3.2L)
- ☒ Crossfire SRT6 (3.2L)
- ☒ MR2 Spyder
- ☒ Prowler (3.5L)
- ☒ PT Cruiser (2.4L)
- ☒ Sebring (2.4L)
- ☒ Sebring (2.5L)
- ☒ Sebring (2.7L)
- ☒ Sebring (3.5L)
- ☒ Sebring Coupe (2.4L)
- ☒ Sebring Coupe (2.5L)

\*\* Sample Models





# Appendix – Vehicle Type

## Pickup Models

- ☒ (Select All)
- ☒ 1500 (3.6L)
- ☒ 1500 (3.7L)
- ☒ 1500 (4.7L)
- ☒ 1500 (5.7L)
- ☒ 1500 Diesel (3.0L)
- ☒ 2500 (5.7L)
- ☒ 2500 (6.4L)
- ☒ 2500 Diesel (6.7L)
- ☒ 3500 (5.7L)
- ☒ 3500 (6.4L)
- ☒ 3500 Chassis Cab Diesel (6.7L)
- ☒ 3500 Diesel (6.7L)
- ☒ Dakota (3.7L)
- ☒ Dakota (4.7L)
- ☒ T100
- ☒ Tacoma
- ☒ Tundra (3.4L)
- ☒ Tundra (4.0L)
- ☒ Tundra (4.6L)
- ☒ Tundra (4.7L)
- ☒ Tundra (5.7L)

\*\* Sample Models





# Appendix – Vehicle Type

## Sedan Models

- ✓ 200 (2.4L)
- ✓ 200 (2.4L) 4 Spd
- ✓ 200 (2.4L) 6 Spd
- ✓ 200 (3.6L)
- ✓ 200 (3.6L) 4 Spd
- ✓ 200 (3.6L) 6 Spd
- ✓ 300 (2.7L)
- ✓ 300 (3.5L) 4 Spd
- ✓ 300 (3.5L) 5 Spd
- ✓ 300 (3.6L)
- ✓ 300 (5.7L)
- ✓ 300 (6.1L)
- ✓ 300 (6.4L)
- ✓ 300 SRT (6.1L)
- ✓ 300 SRT (6.4L)
- ✓ 300M (3.5L)
- ✓ Avalon
- ✓ Avalon Hybrid
- ✓ Camry
- ✓ Camry Hybrid
- ✓ Concorde (2.7L)
- ✓ Concorde (3.2L)
- ✓ Concorde (3.3L)
- ✓ Concorde (3.5L)
- ✓ Corolla (1.6L)
- ✓ Corolla (1.8L)
- ✓ Corolla (2.4L)
- ✓ ECHO
- ✓ LHS (3.5L)
- ✓ Matrix (1.8L)
- ✓ Matrix (2.4L)

\*\* Sample Models



# Appendix – Vehicle Type

## Minivan Models

- ✓ Pacifica (3.6L)
- ✓ Town & Country (3.3L)
- ✓ Town & Country (3.6L)
- ✓ Town & Country (3.8L)
- ✓ Town & Country (4.0L)
- ✓ Town & Country LWB (3.3L)
- ✓ Town & Country LWB (3.8L)

\*\* Sample Models



# Appendix - Targeting Costs

retained_flag	0.000000
predLossProb	0.559311
predClass	0.000000
cumNumOfCustomer	30653.000000
churnIndicator	1.000000
cumChurn	25591.000000
propOfCustomer	0.871418
propOfChurnCaptured	0.919349
lift	1.055003
profit/loss	54.670281
netProfit/netLoss	204912.602237