PREDICTING CARRIER LOAD CANCELLATIONS

MIT Center for Transportation & Logistics

Research Fest

May 22, 2018

AUTHORS

Ali Al-Habib   Nicolas Favier

ADVISOR Dr. Christopher Mejia

AGENDA

1 **INTRODUCTION**
Trucking industry background & load cancellation impacts

2 **DATA ANALYSIS**
Descriptive analytics of load cancellation over three-year dataset

3 **MODELING**
Predictive models applied on the dataset to identify main cancellation drivers

4 **RESULTS**
Models results presented in confusion matrices and results analysis

5 **CONCLUSION**
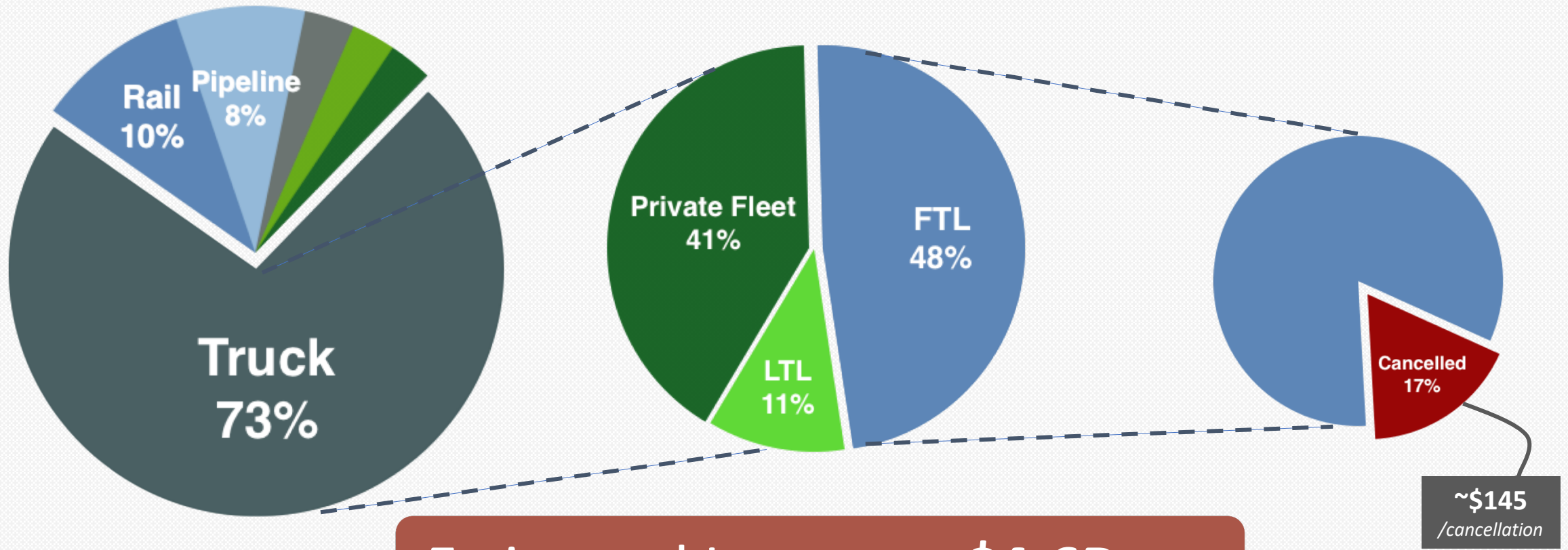Recommended actions and future research challenges

1 INTRODUCTION

MOTIVATION

400 Million Truckloads → **185** Million FTL Truckloads → **32** Million Cancellations

Rail 10%
Pipeline 8%
Truck 73%

Private Fleet 41%
FTL 48%
LTL 11%

Cancelled 17%

~$145 /cancellation

Estimated Impact ≅ $4.6B /year

PROCESS

3-YEAR Dataset
of Full Truckloads

3.6M Records of Full Truckload during 2015, 2016, 2017

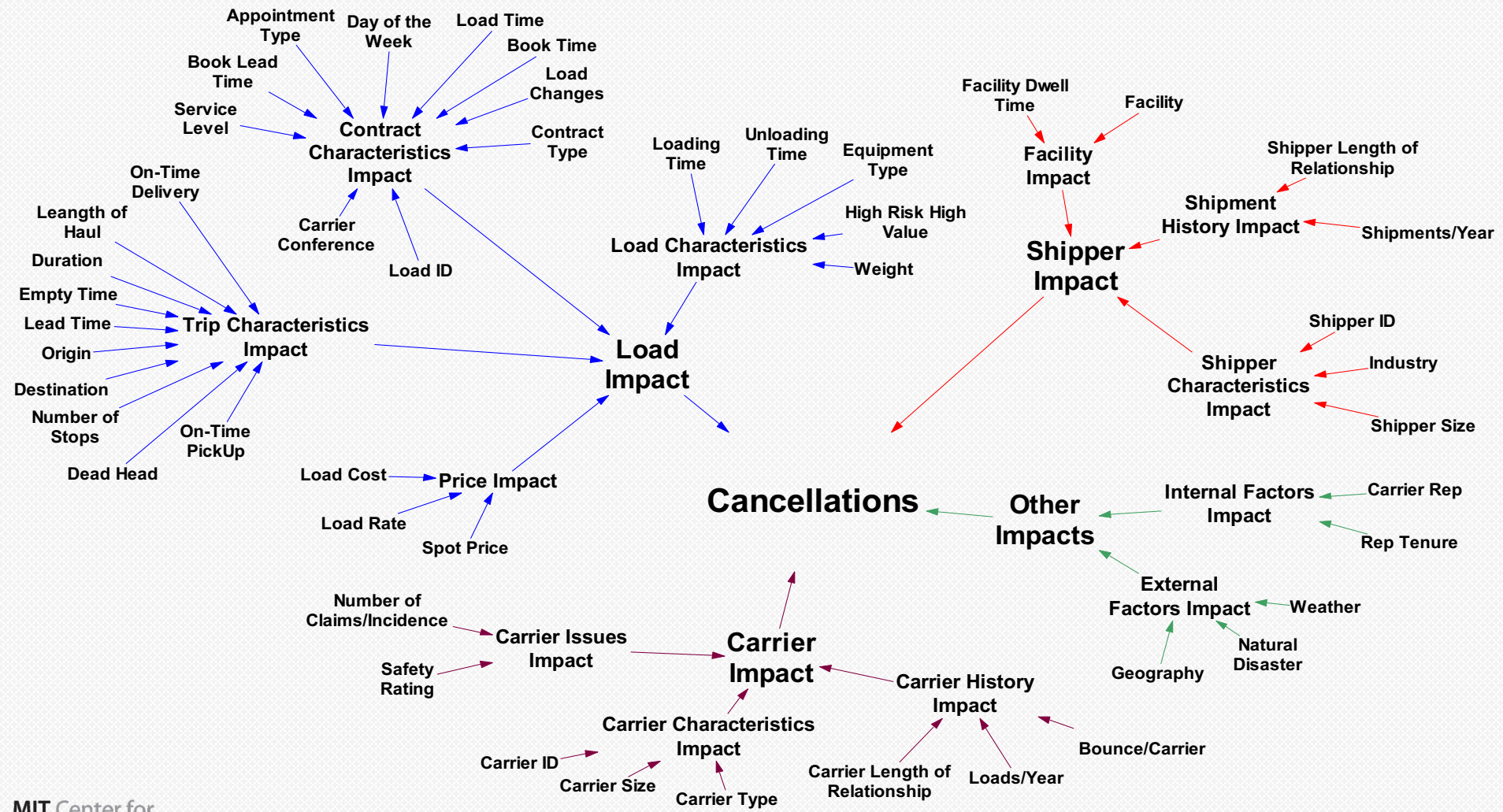## Main Drivers
for Truckload Cancellations

Descriptive analytics to identify the main cancellation drivers

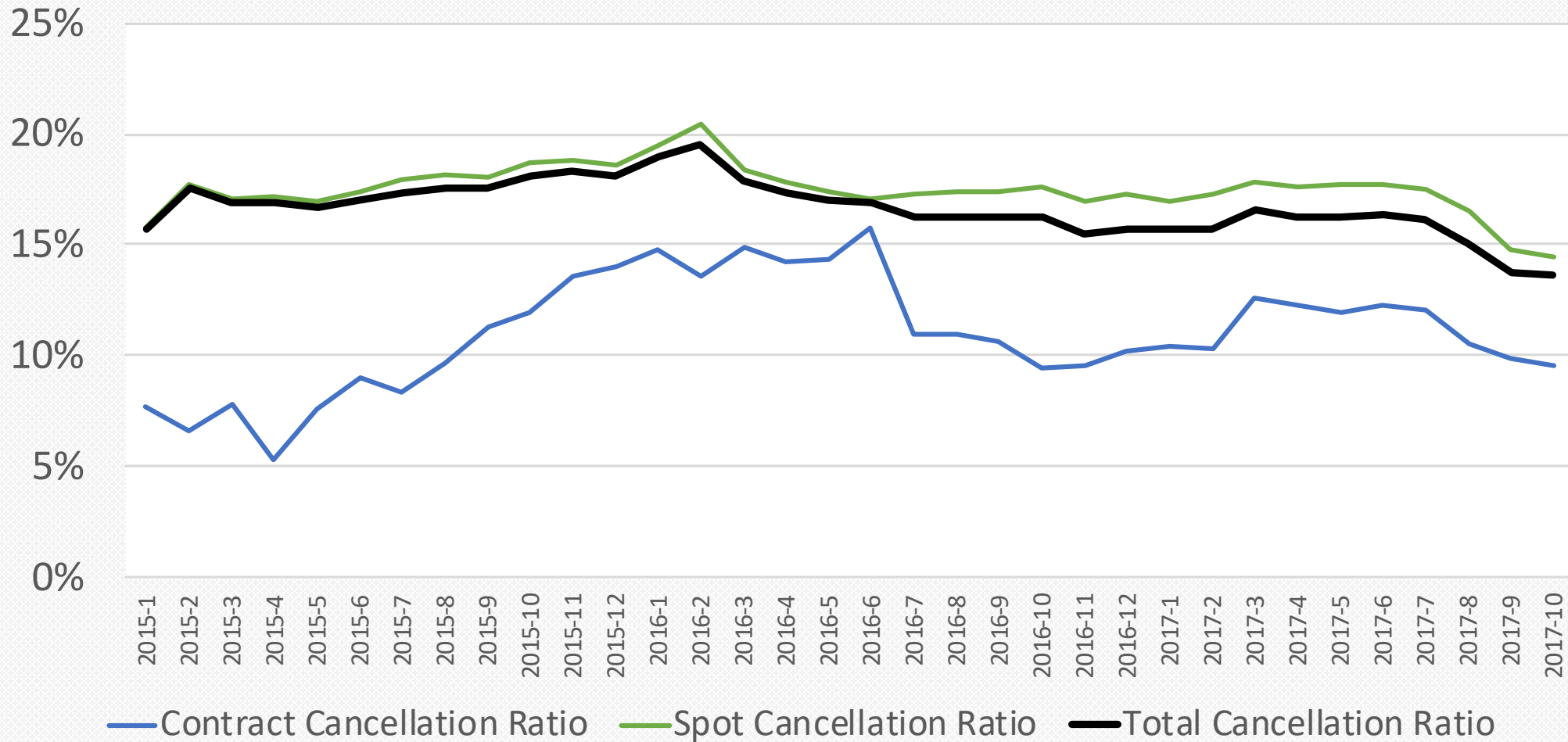## Predictive Model
to Predict Cancellation Probability

Evaluating different models to predict future loads cancellations
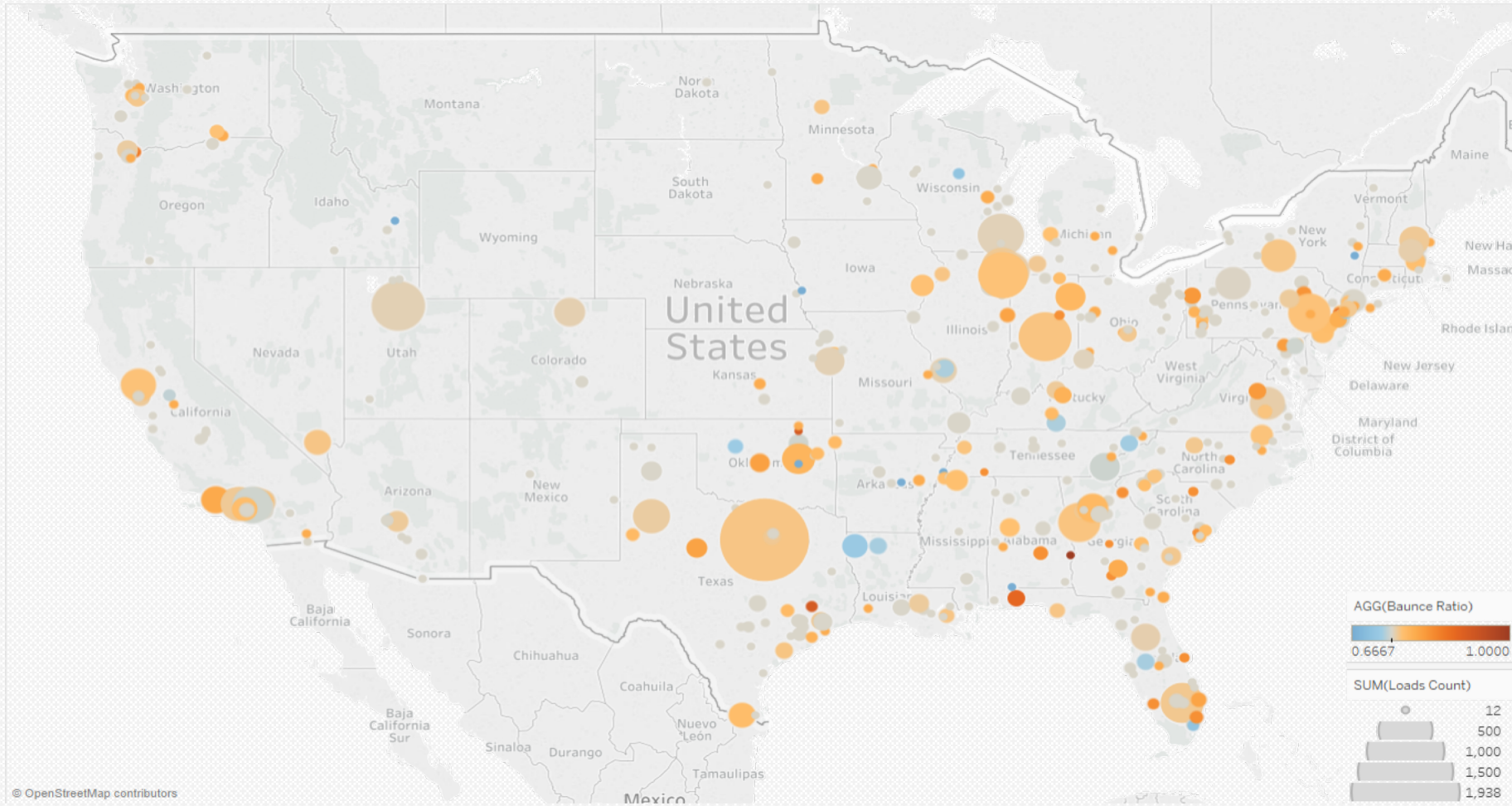
POTENTIAL CANCELLATION DRIVERS

2 DATA ANALYSIS

BEHAVIOR OVER TIME

Cancellation Ratios over time

— Contract Cancellation Ratio — Spot Cancellation Ratio — Total Cancellation Ratio
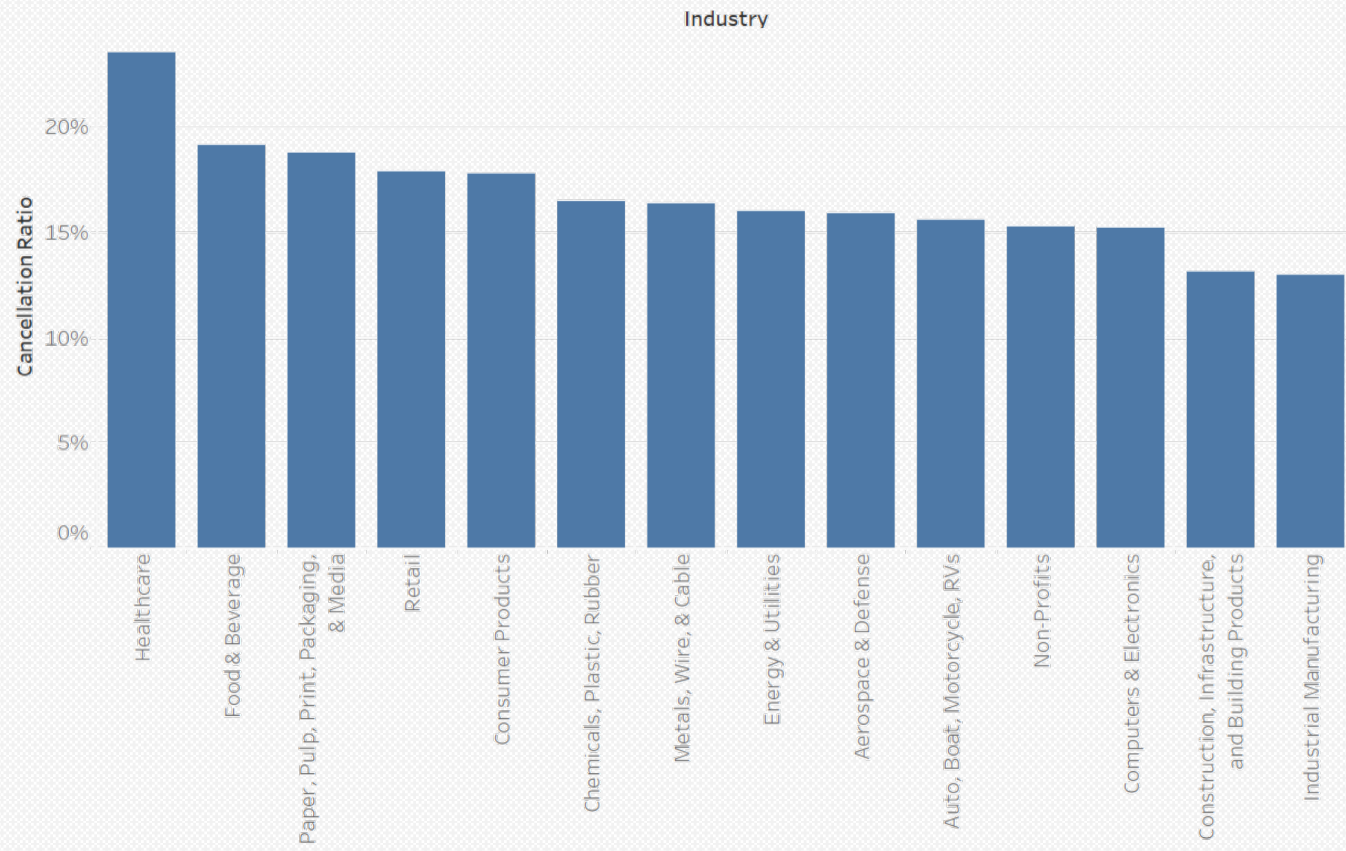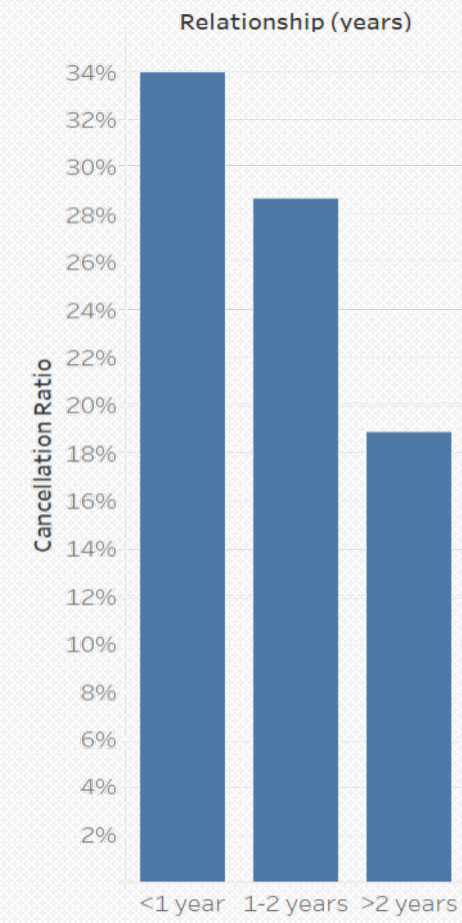
LOCATION FACTOR

Loads & Cancellation Ratios by city
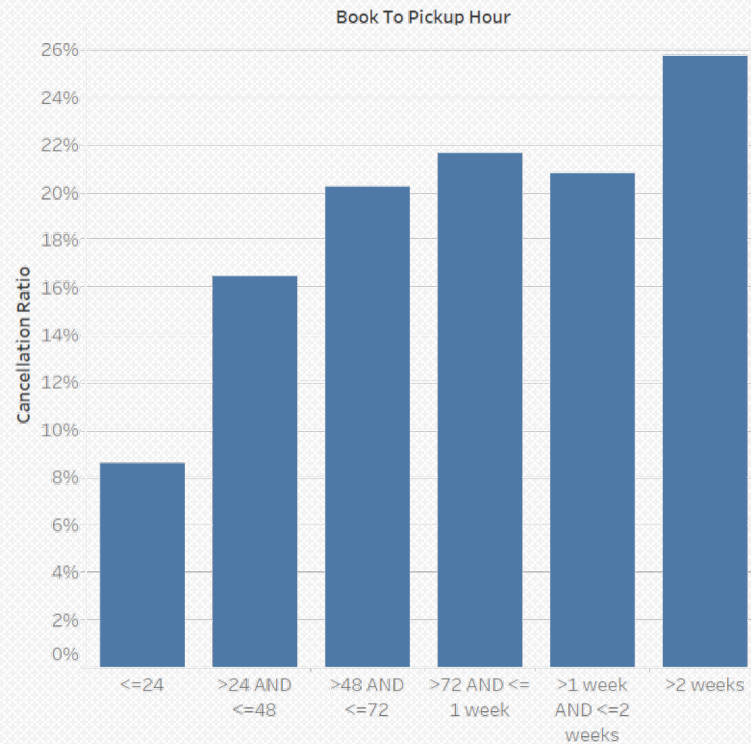
SHIPPERS & CARRIER FACTORS

Cancellation Ratios by shipper industry

Cancellation Ratios by carrier length of relation with the company

TIME FACTORS

Cancellation Ratios by duration between booking & load pickup
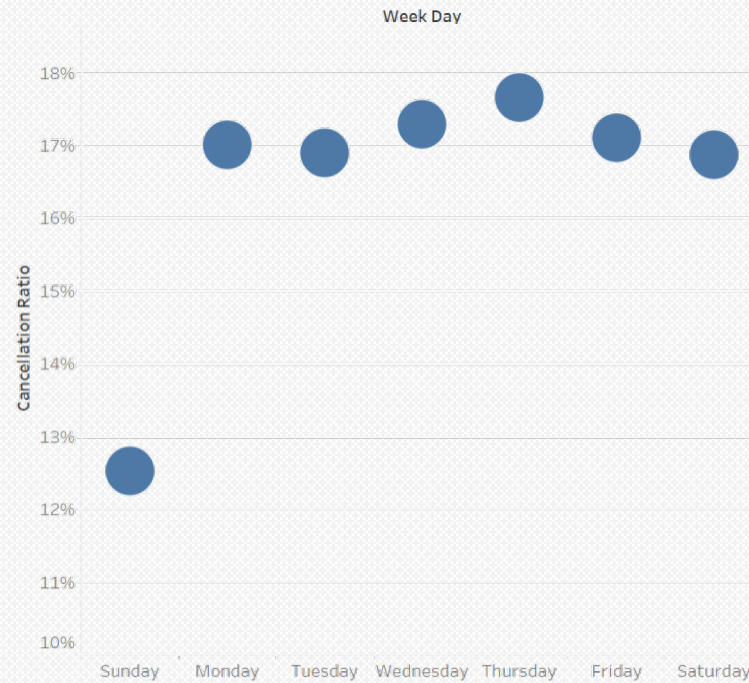
Cancellation Ratios by day of the week
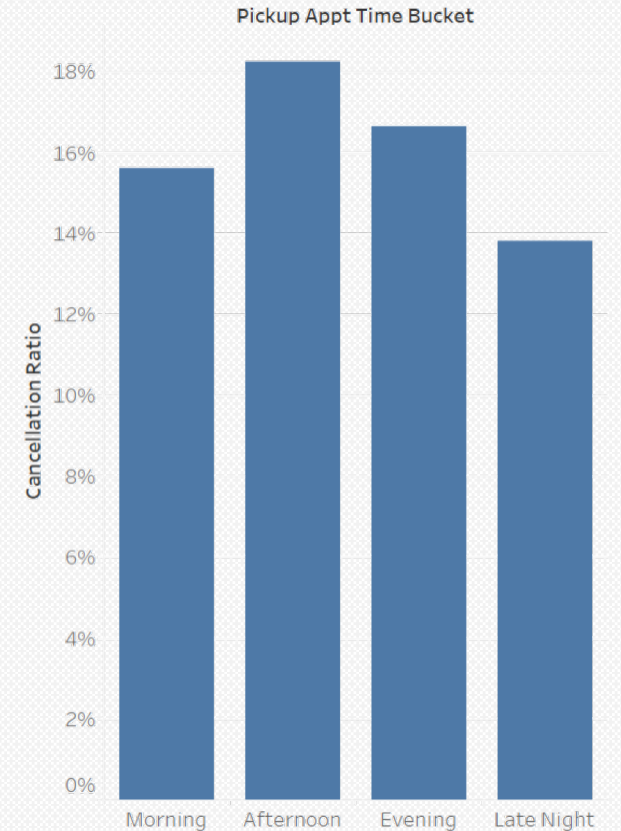
Cancellation Ratios by pickup time

11

3 MODELING

DATA PREPARATION

1
Load-Level Data
Convert data from stop to load level data

2
Outliers Processing
Remove outlier records to avoid undesired impact

3
Correlation
Remove correlated attributes using Correlation & Multi-Collinearity Analysis

4
Predictor Screening
Identify the most significant predictors in the data

5
Build the Model
Build multiple models to predict cancellations & assess results
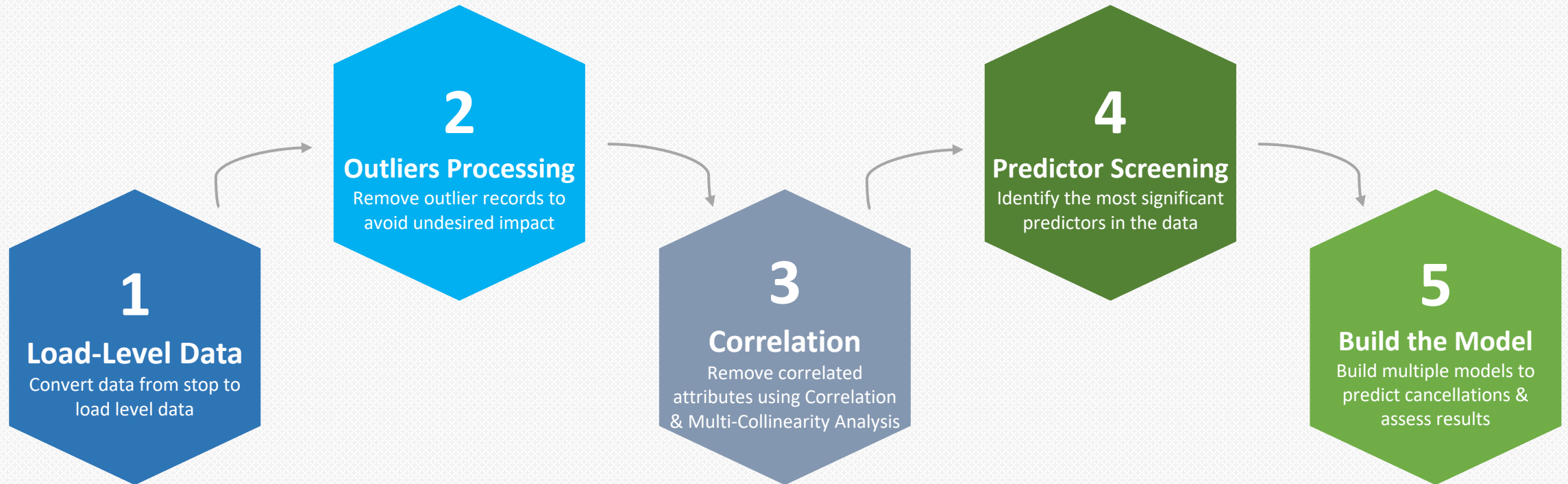
MODELING

LOGISTIC REGRESSION

$$\mathrm{logit}(P) = \ln\left(\frac{P}{1-P}\right)$$
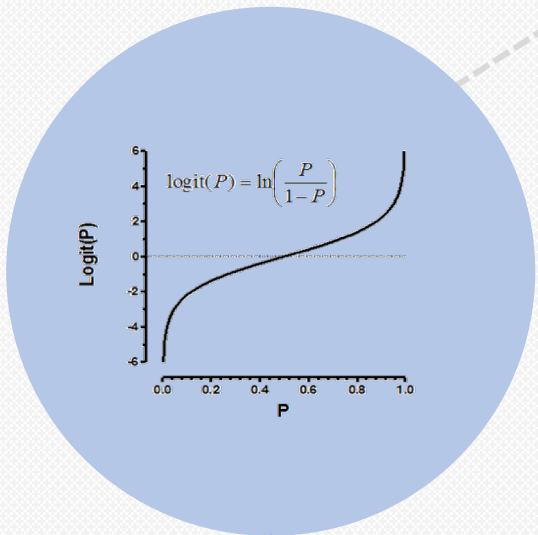
- Categorical Output
- Self-Explanatory
  ▼
- Used as Main Model

MACHINE LEARNING
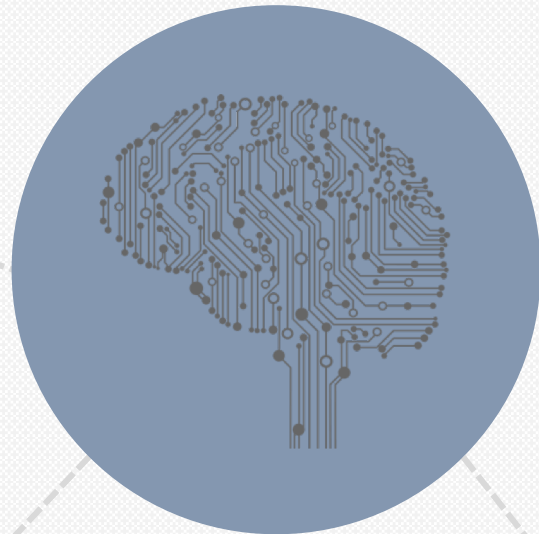
- Multiple Algorithms
- Harder to Explain
  ▼
- Used to Validate Logistic Regression Results
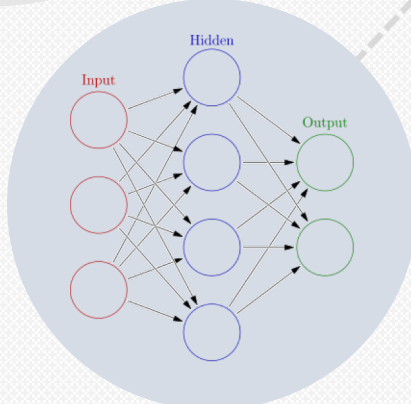
NEURAL NETWORKS
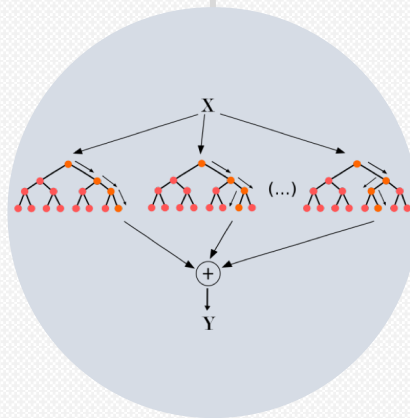
RANDOM FOREST

K-NEAREST NEIGHBOR

4 RESULTS

AVAILABLE DATASET

PREDICTOR SCREENING

Predictor Screening

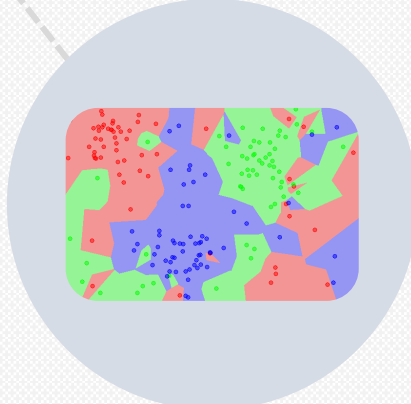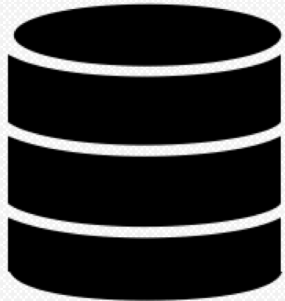| Predictor	IsBounced Contribution	Portion		Rank
BookToPickupHours	46322.8	0.3314		1
BookToLoadDays	39951.2	0.2858		2
LoadVolume	12314.5	0.0881		3
DeadHeadMiles	6417.1	0.0459		4
EquipmentPowerUnits	4513.7	0.0323		5
carrier_Contract-Spot	4512.8	0.0323		6
EquipmentDrivers	4180.3	0.0299		7
MaxDwellTime	2656.2	0.0190		8
Cost	2492.0	0.0178		9
Contract-Spot	2341.5	0.0168		10
MarketCost	2249.9	0.0161		11
PickupState	2079.6	0.0149		12
Industry	1432.3	0.0102		13

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	2.20484041	0.2119005	108.27	<.0001*
DeadHeadMiles	-0.0009195	2.0671e-5	1978.7	<.0001*
BookToPickupHours	-0.0060589	2.1234e-5	81418	<.0001*
EquipmentPowerUnits	-4.4072e-5	2.2362e-6	388.42	<.0001*
LoadVolume	0.00011912	9.0133e-7	17465	<.0001*

MODEL RESULTS

		Predictions		
		No	Yes	
Actual	No	652,501	2,956	**655,457**
	Yes	129,727	1,971	**131,698**
		782,228	**4,927**	**787,155**
Error				16.86%
Missed Bounces				98.50%

	Error %	Missed Bounces
Neural Networks	16.73%	99.95%
Random Forest	16.61%	99.48%
K-Neares Neighbor	19.90%	84.44%

AVAILABLE DATASET

16

DATA ENRICHMENT

CANCELLATION RATIOS

Loadid	CarrierID	Sequence	CityName	IsBounced
4997327	80887	1	Rochelle	False
5031474	80887	1	Rochelle	False
6147652	80887	1	Rochelle	False
6268918	80887	1	Rochelle	False
8370937	80887	1	Rochelle	False
7925070	80887	1	Rochelle	False
8310781	80887	2	Rochelle	False
8692122	80887	1	Rochelle	False
7876415	80887	1	Rochelle	True
7925070	80887	2	Rochelle	False
8107515	80887	1	Rochelle	False
8308590	80887	2	Rochelle	False
9202460	80887	1	Rochelle	False
9202460	80887	2	Rochelle	False

Carrier (80887) & City (Rochelle)
Bounce Ratio=1/12=**0.08333**

Repeated loads are counted only once for the ratio calculation

Loadid	CarrierID	CityName	Sequence	carrierCityBounceRatio
8370937	80887	Rochelle	1	0.083333
8370937	80887	Wintersville	2	0.222222

Average of the CarrierCity Bounce Ratio for Each Stop

Loadid	CarrierID	carrierCityBounceRatio
8370937	80887	0.152778

Aggregated carrierCityBounce Ratio on Load-Level
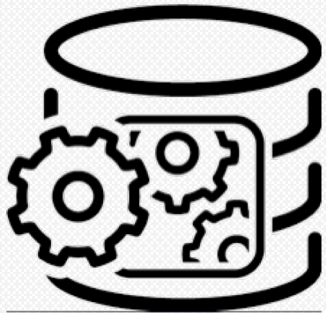
SEVERE WEATHER DATA*

ENRICHED DATASET

17

ENRICHED DATASET

PREDICTOR SCREENING

Predictor Screening

Predictor	IsBounced Contribution	Portion		Rank
carrierCityBounceRatio	733035	0.8383		1
CarrierEquipmentTypeBounceRatio	53397	0.0611		2
BookToPickupHours	38846	0.0444		3
ShipperBounceRatio	15532	0.0178		4
LoadVolume	13612	0.0156		5
DeadHeadMiles	3163	0.0036		6
EquipmentPowerUnits	2774	0.0032		7
Contract-Spot	2009	0.0023		8
carrier_Contract-Spot	1923	0.0022		9
PickupState	1819	0.0021		10
DropState	1293	0.0015		11
Miles	1038	0.0012		12
MaxDwellTime	920	0.0011		13

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	4.58208581	0.009991	210332	<.0001*
carrierCityBounceRatio	-11.960166	0.0188223	403767	<.0001*
CarrierEquipmentTypeBounceRatio	3.31025278	0.0284327	13555	<.0001*
BookToPickupHours	-0.005077	2.8018e-5	32834	<.0001*
ShipperBounceRatio	-3.0534413	0.0493792	3823.8	<.0001*
Contract-Spot[C]	-0.3171286	0.0029911	11241	<.0001*

MODEL RESULTS

		Predictions		
		No	Yes	
Actual	No	638,652	16,880	655,532
	Yes	52,155	79,468	131,623
		690,807	96,348	787,155
Error				8.77%
Missed Bounces				39.62%

	Error %	Missed Bounces
Neural Networks	8.67%	39.04%
Random Forest	8.70%	42.13%
K-Neares Neighbor	9.33%	44.32%

ENRICHED DATASET

18

ADDITIONAL DATASET

| Dataset (~3-year data) | | Additional 3-month data |

| Cancellation Ratios Calculation (100%) | → | Ratios |

| Training (80%) | Testing (20%) |
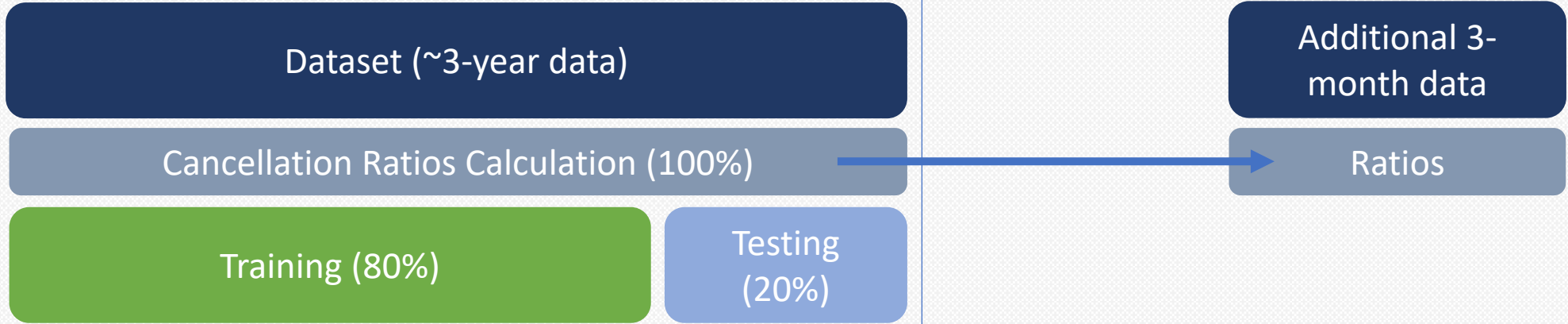
ENRICHED DATASET

Predictions

Actual		No	Yes	
	No	638,652	16,880	**655,532**
	Yes	52,155	79,468	**131,623**
		690,807	**96,348**	**787,155**
Error				8.77%
Missed Bounces				39.62%

NEW DATASET

Predictions

Actual		No	Yes	
	No	59,883	3,735	**63,618**
	Yes	8,903	1,722	**10,625**
		68,786	**5,457**	**74,243**
Error				17.02%
Missed Bounces				83.79%

	Error %	Missed Bounces
Neural Networks	16.78%	84.70%
Random Forest	16.19%	87.98%
K-Neares Neighbor	16.41%	86.66%

UNPREDICTABILITY TESTING

AVAILABLE HISTORICAL DATA

Additional 3-month data	
<= 10 Historical Records (67%)	> 10 Historical Records (33%)

Predictions

		No	Yes	
Actual	No	21,449	368	**21,817**
	Yes	2,222	542	**2,764**
		23,671	**910**	**24,581**
Error			10.54%	
Missed Bounces			80.39%	

PREDICTION TIME HORIZON

Additional 3-month data	
<= 10 Historical Records (67%)	> 10 Historical Records (33%)
	7-day Horizon (3%)

Predictions

		No	Yes	
Actual	No	2,147	31	**2,178**
	Yes	176	44	**220**
		2,323	**75**	**2,398**
Error			8.63%	
Missed Bounces			80.00%	

MULTIPLE CLUSTERS, MULTIPLE MODELS

Test		Error	Missed Bounces
Logistic Regression (Threshold=0.5) - Base Scenario		17.02%	83.79%
Cost Clustering	Low Cost (<= $500)	18.20%	99.06%
	Mid Cost	16.67%	98.46%
	High Cost (>= $6000)	8.49%	100.00%
Miles Clustering	Same day delivery (<= 250 mi)	16.07%	99.18%
	Next Day delivery	18.08%	98.18%
	Long Haul (>= 550 mi)	18.08%	98.18%
Book To pickup Hours Clustering	Less than 24h	8.53%	100.00%
	Between 24h and 48h	16.91%	100.00%
	Between 48h and 72h	20.58%	99.99%
	More than 72h	22.33%	99.58%

THRESHOLD SENSITIVITY ANALYSIS
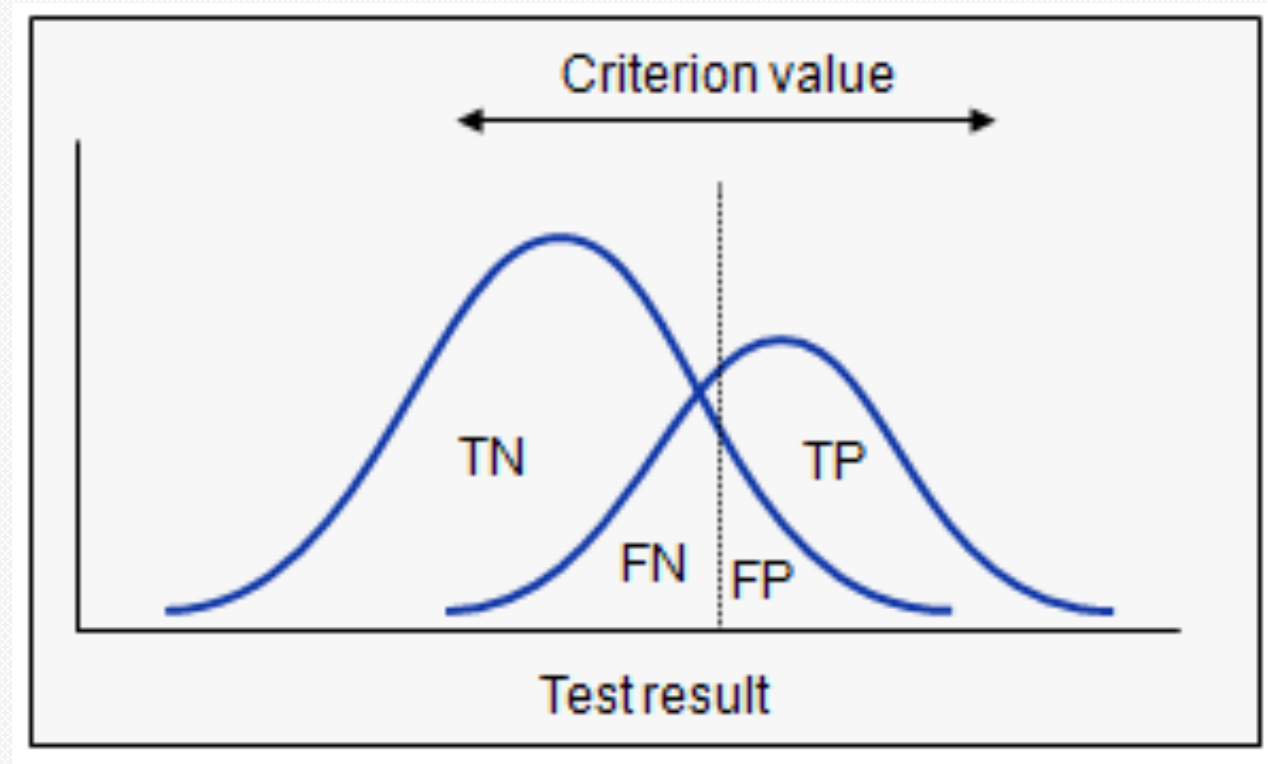
THRESHOLD SENSITIVITY ANALYSIS
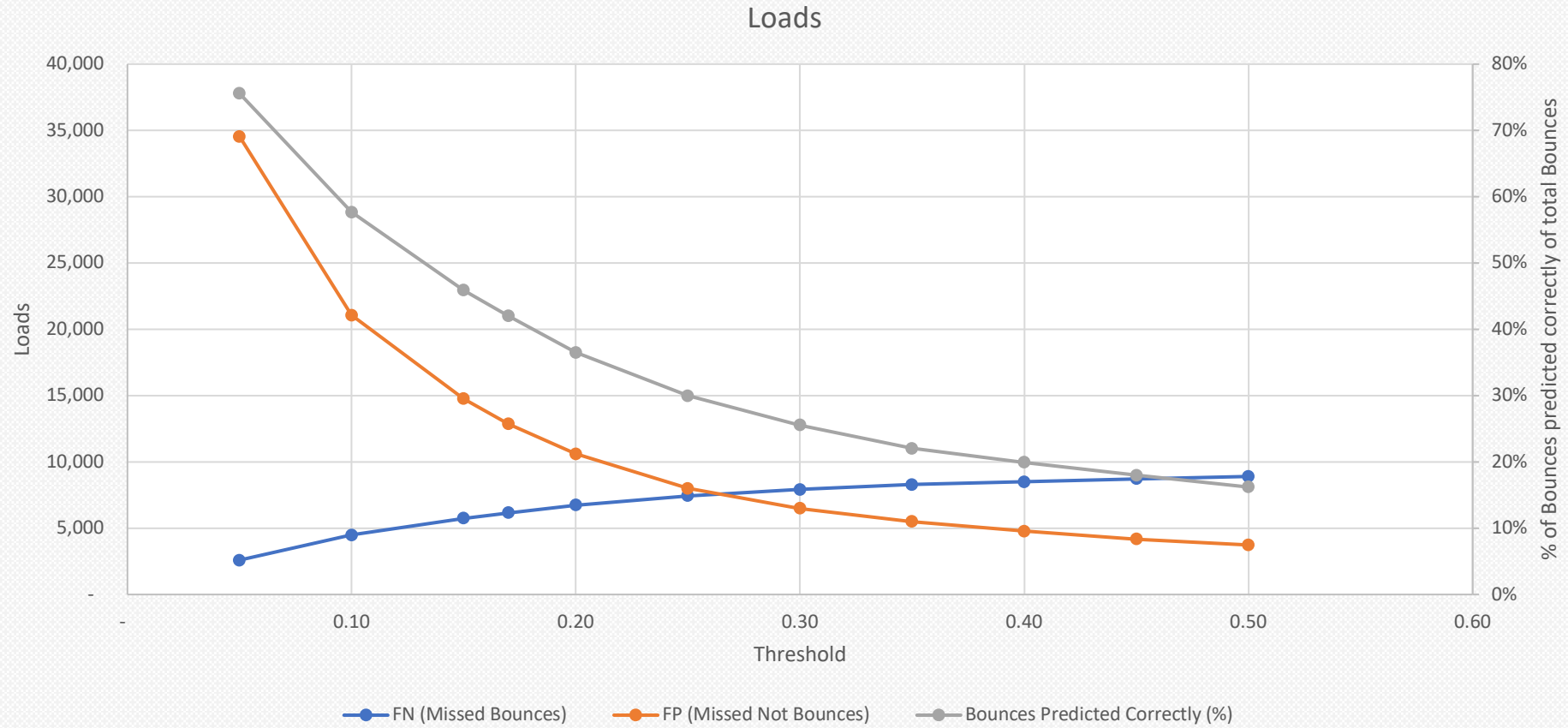
Loads

5 CONCLUSION

NEXT STEPS

THRESHOLD CHANGE

- Use the model with lower threshold (0.17)
- Predict up-to **42%** of cancelled loads
- Tradeoff ratio **4:1**
 - (predicted cancellation : actual cancellation)

FURTHER RESEARCH

- Surveys to capture range of cancellation reasons
- Record actual reasons for each cancellation
- Capture details related to these reasons
- Record additional information for each load:
  - Loads sequence at truck level
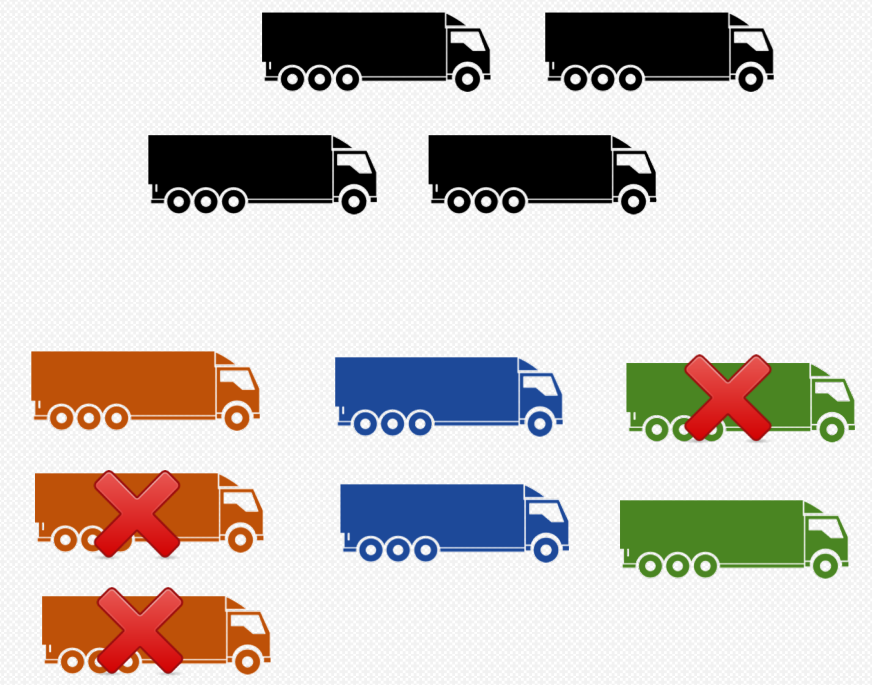 - Carrier booked capacity
  - Rejection Rate

CHALLENGES

LOAD SEQUENCE SCENARIO

OVERBOOKING SCENARIO
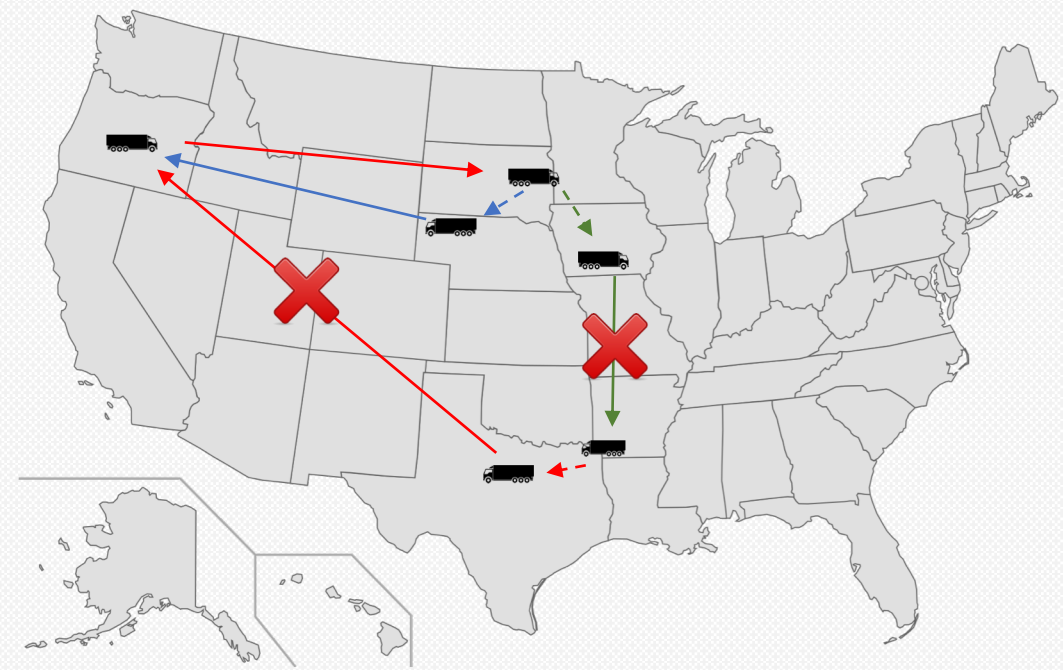
- 🔴 COMPANY A
- 🟢 COMPANY B
- 🔵 COMPANY C
- 🟡 SELECTED ROUTE

26