

Swire Coca-Cola Business Churn Prediction



Cole Lifer

Jordan Harmer

Richard Garbett



**David Eccles
School of Business**

THE UNIVERSITY OF UTAH

Problem Statement

What is Swire looking for?

- **Profitable investment**
- **Longevity**
- **Loyalty**

What Can They Offer?

- **Discounted pricing for loyalty**
- **Various Incentives**

What are the Risks?

- **Incentivizing unprofitable customers**
- **Equipment Loss**

Dataset

Problems

- Onboarding date
- Outliers in dates
- Unclear target variable

Positives

- Diverse dataset
- Few missing values
- Categorical variables

Exploratory Data Analysis

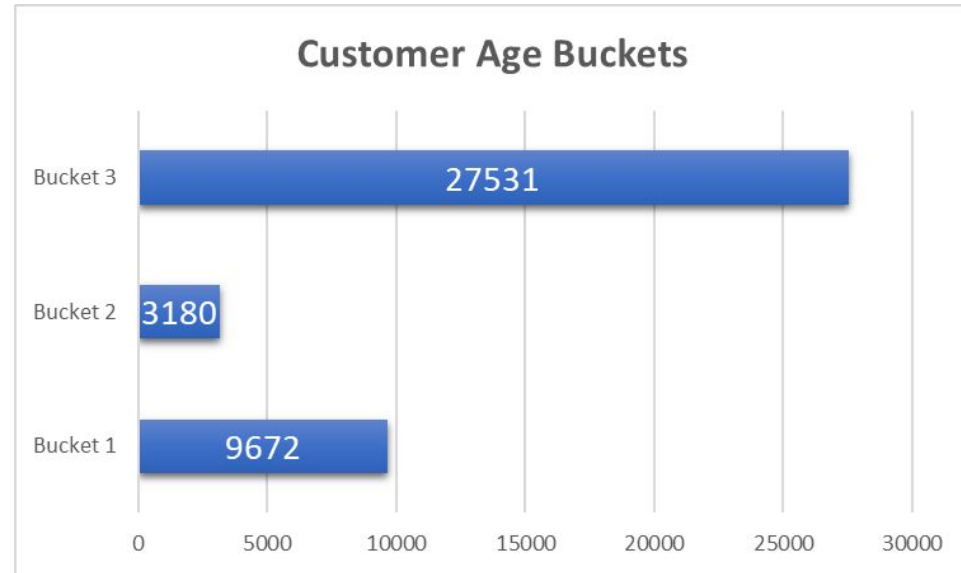
Transaction Size and Customer Age

- Average Customer Age: **8.8 years**
- Invoice Range: **\$-90.95** to **\$1.8M**
- Transaction Count Range: **1** to **468**

Exploratory Data Analysis (Continued)

Customer Age Buckets

- 2 years or newer = **Bucket 1**
- 2- 3 years old = **Bucket 2**
- 3+ years old = **Bucket 3**



Initial Model Development

Random Forest

- Inherent feature selection
- Handles large volume of predictors
 - More prone to overfitting
(Far less than single decision tree)
 - Compute intensive
- Imbalanced datasets = problematic

Naive Bayes

- Efficient
- Better for categorical data
- Not optimal for rare situations
(Pros & Cons)
- Baked-in assumptions
(variable independence)

Model Selection



Initial Performance

- Admirable Performance from Both
-



Iterative Process

- Evaluate, Process, Edit, Re-evaluate

Model Iteration

➤ Variable Reduction

- Chi-squared test
- Overlapping areas

➤ Downsampling Strategies

➤ 60/40 Train/Test Split

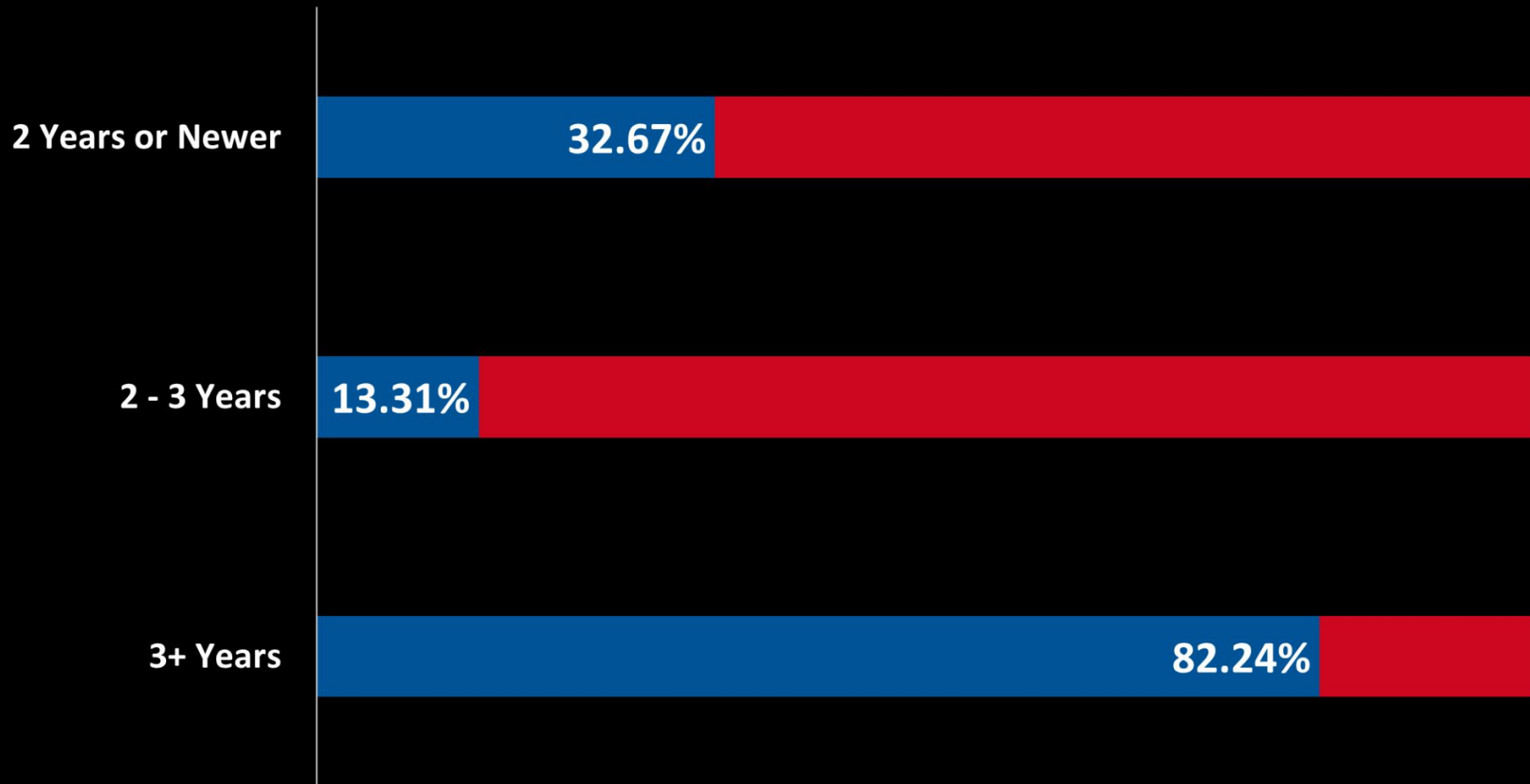
➤ 3-fold cross validation

Performance Metrics

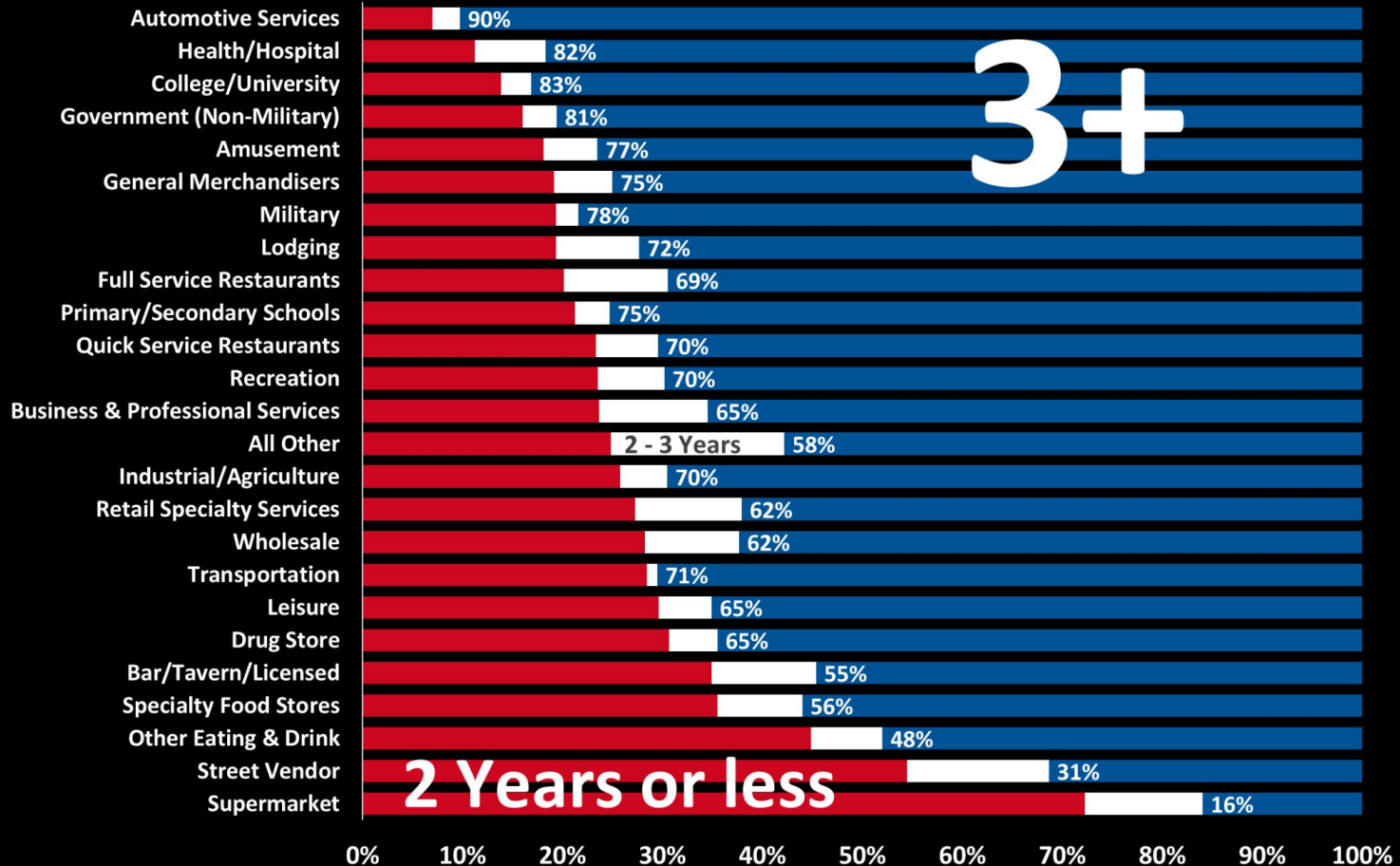
- 67% Mean Accuracy
- Best Proficiency in Third Bucket

Titles	Fold 1	Fold 2	Fold 3	Mean	Std. Dev.
Accuracy	67.25	67.72	67.11	67.36	0.32
Precision (1)	41.45	43.09	42.22	42.25	0.82
Precision (2)	30.8	31.37	36.4	32.85	3.09
Precision (3)	72.72	72.58	72.6	72.63	0.07
Recall (1)	25.25	24.57	28.29	26.03	1.98
Recall (2)	8.4	8.02	7.83	8.08	0.29
Recall (3)	88.81	89.78	87.6	88.73	1.09
F-score (1)	31.38	31.29	33.88	32.18	1.47
F-score (2)	13.19	12.77	12.89	12.95	0.22
F-score (3)	79.96	80.27	79.4	79.88	0.40

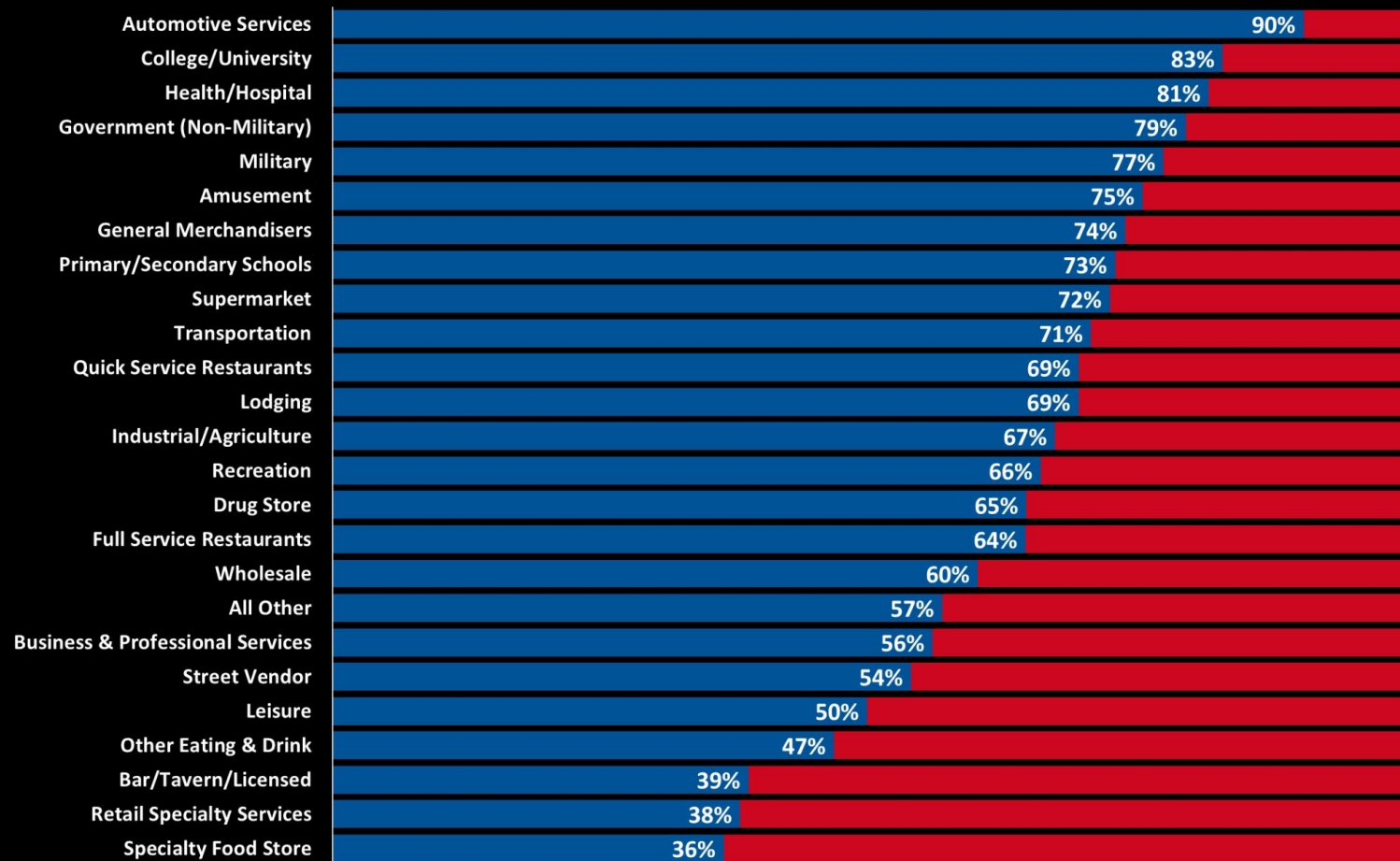
Prediction Accuracy by Customer Age Bucket



Customer Age Bucket by Segment



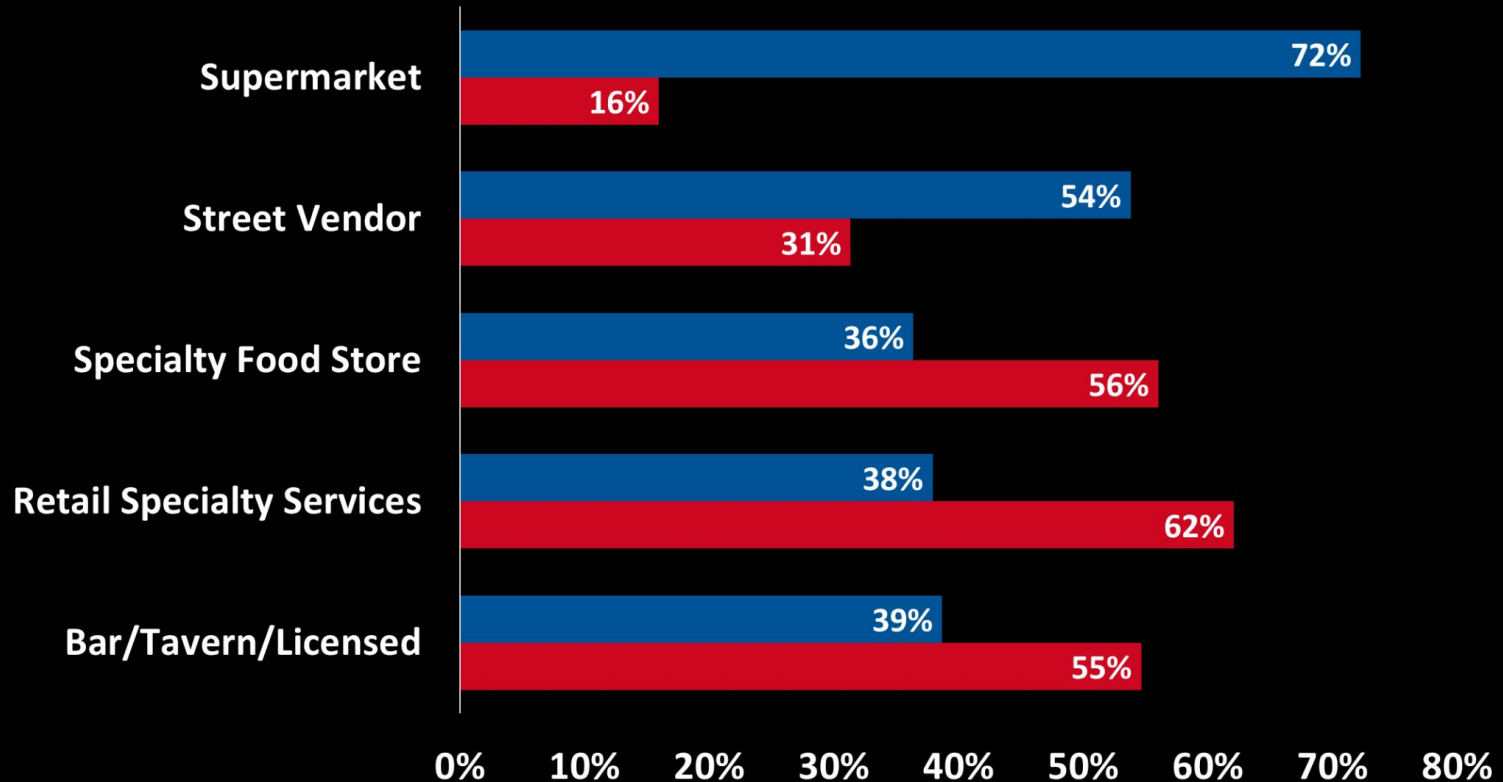
Prediction Accuracy by Segment



Prediction Accuracy

vs.

Share of 3+ Year Old Customers





Next Steps

