

## Data Science Case Study

# Customer Churn Risk Scoring Using Machine Learning

*Mosaic utilized different machine learning approaches to help this retail energy company combat customer churn. [This case study builds off segmentation work Mosaic performed for the same customer.](#)*



### Industry

[Retail & Propane](#)



### Use Case

Customer Churn Prediction



### Techniques

[Classification & Adoption & Scale](#)



### Outcome

Targeted Retention Efforts, Less Churn, and Market Intelligence

# THE CASE FOR IDENTIFYING CUSTOMER CHURN & PROMOTING CUSTOMER RETENTION

Retaining customers is a must for a company's bottom line. A company's customers are its greatest asset, impacting business now, and becoming more valuable over time as they continue to invest in products and services. Customer churn can be costly, or even devastating, to growing and established organizations alike. The true cost of churn is often higher than business leaders generally estimate. Not only does it lead to lost revenue in the near term, but it also means your team must double down on acquiring new customers to fill those revenue streams to ensure continued success in the future. It is widely accepted that it can cost up to 5 times as much to acquire a new customer as it does to retain a current customer.

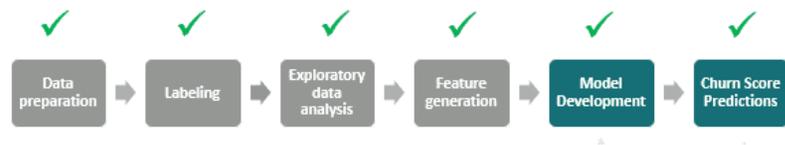
Mosaic's client, a leader in the propane industry, had seen a sharp rise in customer attrition. Recognizing the implications to their business in having to win new customers at a steep cost, they wanted to prevent further customer loss by learning why customers were terminating service. They turned to Mosaic, a trusted partner who had helped them identify regions to successfully target new customers using unsupervised learning techniques. Based on our prior work on this customer segmentation project, Mosaic was tasked with proving the value of applying machine learning to combat customer churn.

# CHURN PREDICTION

Mosaic leveraged historical data used in a previous project and used real examples of customers deciding to leave to learn the attributes and behavior that typically precede customer turnover. Mosaic's data scientists interviewed subject matter experts to incorporate their expertise in developing a working ML model.

Data-driven organizations like the propane firm typically use customer segmentation as a foundation for other value-added analytics. Customer churn is a natural next step since it leverages the knowledge and data of the customer segmentation project. Churn prediction enables targeted marketing and direct intervention for customers most likely to leave, streamlining use of the marketing budget.

## MOSAIC'S ML APPROACH



### Feature Engineering

For any ML approach to be effective, substantial effort must be made to engineer the data in a format that makes sense. For this effort, Mosaic developed a feature table describing what was occurring with any given customer in any given month. Next, the team imputed null values. The team constructed features



using time parameters to build features that look back in a customer's history before they decide to leave. To standardize the different paths a customer might take in their history, features were placed on a relative scale rather than using absolute numbers.

## Algorithm Selection

When the data was ready for analysis, Mosaic evaluated the following three classification algorithms for performance and ranking suitability:

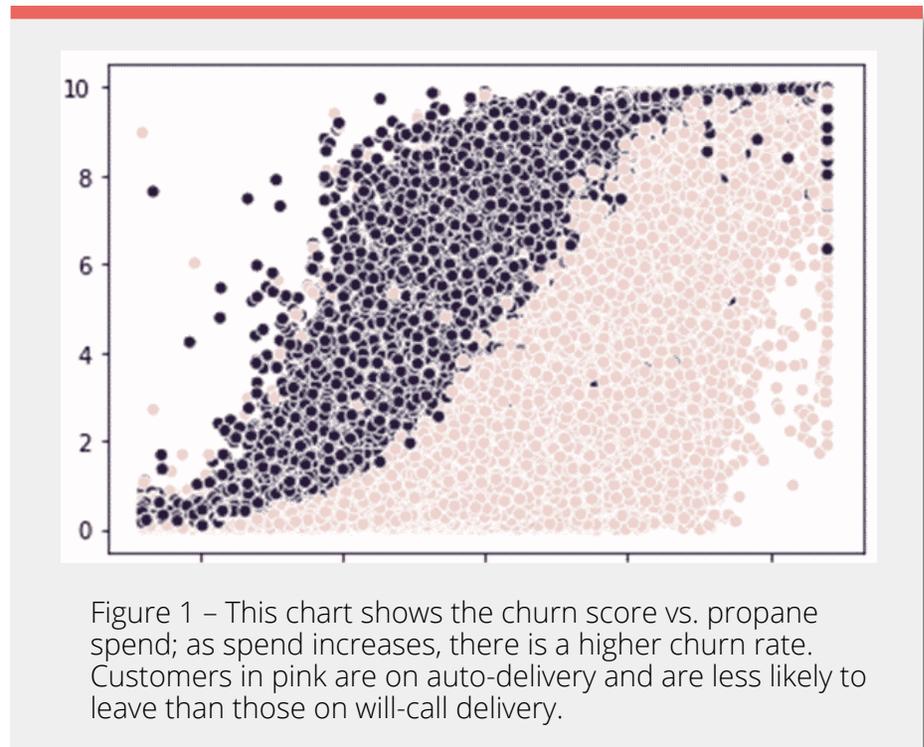
- Simple Decision Tree
- Random Forest (ensemble of decision trees)
- Logistic Regression

In training runs, the random forest and logistic regression (logit) algorithms had similar performance, but due to precision & recall metrics, logit was determined to be more suitable. Logit enables more fine-grained predictions than tree-based algorithms; in other words, there is much less chance two customers have the same churn score.

## Churn Model Performance

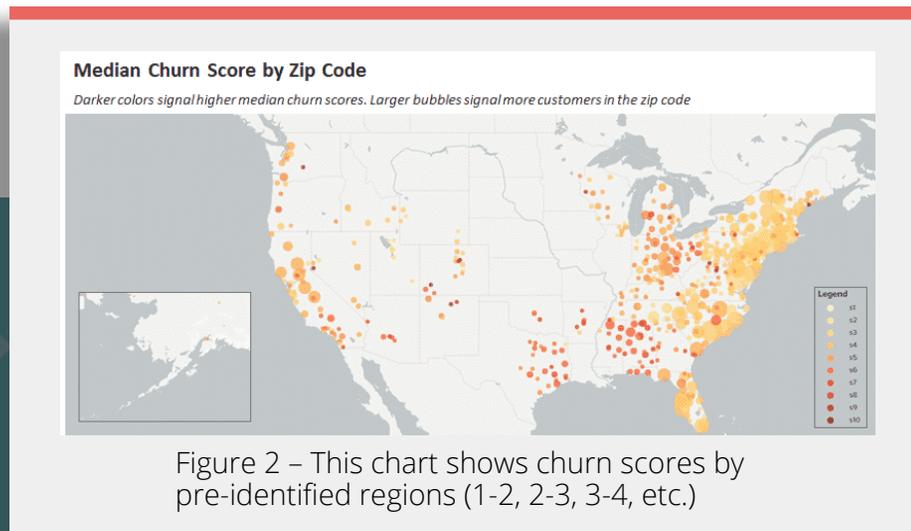
Mosaic's data science consultants were able to develop a fine-tuned churn model using the Logit algorithm. Upon validation, the logit model was able to predict churn ~80% accurately. Logit allowed the team to use all variables related to a customer's account with the propane firm, rather than being limited to a handful of top features.

The model generated a rank-ordered list of churn scores for all customers using the latest data available. A low score means a customer is less likely to leave, the higher the churn score, the more or less likely that the customer will leave. The rank-ordered list of customers can be further drilled down to identify churn drivers by location, account type, and many other attributes a marketing analyst or business leader would want to explore. The model teased out drivers of churn the client had not previously identified, such as delivery method, usage increases/decreases, tank size, and time. Figure 1 shows certain factors that were associated with a higher likelihood that a customer would leave, such as level of spending and delivery method.



## Getting these insights into decision-makers' hands

These insights are incredibly valuable to the marketing team as they can use this knowledge in developing more effective campaigns and conducting outreach to improve retention. Figure 2 shows churn rates by region, enabling marketers to target customers for specific intervention.



## MOVING FORWARD WITH CONFIDENCE

The propane company is now able to use churn rankings to inform targeted marketing approaches and intervention strategies. These insights can be shared with field operations leaders to establish a data-centric risk management approach. The company can also incorporate the churn score into customer service representative systems. To achieve optimal accuracy, the customer churn model will be refreshed periodically. Mosaic's work on both customer segmentation and churn modeling has yielded a robust Customer Lifetime Value metric the firm can measure moving forward.

