Online Grocery & Omnichannel Strategy

By: Ryan Alexander Alberts & Antoine Lahad Abinader
Advisor: Dr. Eva Ponce

Topic Areas: Distribution, Last Mile, Forecasting

Summary: Groceries present a unique category for eCommerce due to particularly onerous complications from last-mile delivery of fresh products. This study analyzes consumer preferences for omnichannel fulfillment, focusing on two channels: home delivery and pick-up in store. Specifically, it identifies critical markets for home delivery of online grocery and provides insights into drivers of channel choice.

Before coming to MIT, Ryan took over, ran, and sold an eCommerce retailer in the home improvement category. He holds a B.A. from Columbia University.

Before coming to MIT, Antoine implemented and managed warehouse solutions for SPAN. He holds a B.Eng. from the American University of Beirut.

KEY INSIGHTS

1. Online grocery retailers can benefit from data-driven omnichannel strategy.
2. Drivers of channel adoption are different for pickup and home delivery.
3. Retailers can shape demand for particular channels according to channel attributes.

Introduction

Every aspect of the grocery shopping experience is evolving through eCommerce. Channel choice, i.e. the choice between fulfillment options, is evolving as customer expectations outpace retailers' ability to deliver cross-channel experiences.

The MIT team worked with Walmart to study omnichannel strategy in the context of online grocery. The competitive landscape is such that Walmart has made significant efforts to expand the capabilities of its network of retail locations to satisfy shifting consumer habits.

The traditional way to reach customers in e-commerce is home delivery. More recently, retailers have started to offer different options for fulfilling online orders, including picking up from the retail store, picking up from lockers, and picking up from collection points (e.g. convenience stores, gas stations, etc.). This project focuses on two channels: pick-up from store and home delivery.

Launching a home delivery service requires systems, processes, and capabilities that involve significant up-front investment, in addition to network and warehouse design implications for opening new sites. Moreover, the push toward extending coverage of home delivery capabilities is at the forefront of most brick-and-mortar retailers' competitive strategy to combat pure-play eCommerce players. To that end, understanding the online grocery customer is a necessary first step. This research explores customer purchasing behavior and distribution channels features to answer the below driving questions:

1. What are the critical US markets for home delivery?
2. What drives customers channel choice?

Methodology

The MIT team formulated a framework with two pipelines to handle two data sets. The first is the customer profile pipeline, which flows from historical customer data provided by Walmart. Using existing customer purchasing behavior data, the MIT team designed an algorithm based on regions where both channels, home delivery and pick-up, are active. The model was then deployed to predict home delivery adoption in regions that do not currently have home delivery service.

The second is the channel choice pipeline, flowing from a Discrete Choice Experiment (DCE) devised by the MIT team, consisting of a survey of the retailer's customers; the survey responses serve as the basis of study. These responses serve as the input to logistic regression models that capture 1) a customer's channel preference, 2) what features drive that preference, and 3) how sensitive customer preferences are to changes in features. Together, the findings of the two pipelines complement each other in answering the driving questions of the research stated in the introduction.

The MIT team segmented the retailer's online grocery customers by their total number of home deliveries and orders picked up in store. The target variable is home delivery adopters, which order two or more orders fulfilled via home delivery.

Findings – Customer Profile Pipeline

The team first conducted exploratory analysis of the Colorado market, given that it was the largest and most mature online grocery market for Walmart. The MIT team ran a k-means cluster analysis on population density per zip code ('pop_density'), number of online grocery competitors per zip code ('total_comp'), the average value of an online order aggregated by zip code ('mean_sales') and the number of home delivery adopters per zip code. The result was two distinct clusters with clear distinctions between the city center and the suburban and rural areas. The first groups the city area, defined by high population density and number of competitors, as well as low mean sales per customer. Meanwhile, the second cluster covers suburbs where there is low population density, low number of competitors and high mean sales per customer.

Cluster analyses like these allowed the MIT team to explore feature sets that would be useful for predictive models and informed how we built our model.

Several classification models were compared according to the F1 score of their predictions to evaluate which would be used to extrapolate insights from the 5 pilot regions to the larger US market. These models included gradient boosting machines (GBM), k-Nearest Neighbors, and Naïve Bayes.

The GBM model produced the highest F1 score at 0.88, edging out the kNN and Naïve Bayes models, which each produced F1 scores of 0.87. This score effectively means that the GBM model more effectively minimized misclassifications while maximizing accurate classifications. Given that the F1 score captures the trade-off between precision and recall, the score evaluates the model's ability to be both accurate and generalizable, key characteristics of predicative models.

The GBM model F1 score is significant in the context of the null error rate – how often the model would be wrong if it always predicted the majority class, i.e. non-home-delivery – which is 84.36% (8,996 out of 10,663). This null error rate is useful context in that, picked at random, any given observation is far more likely to be a non-home-delivery customer. Thus, accurately predicting 58.66% of customers that were actually home delivery customers is meaningfully higher than random chance (which would yield 15.64% probability of being a true home delivery customer).

The model predictions were aggregated by zip code. The zip codes were ranked by predicted number of home delivery adopters and by predicted density of home delivery adopters.

The model also allowed the MIT team to dive deeper into looking at the top features by feature importance, finding significant trends.

The MIT team performed an Ordinary Least-Squared (OLS) regression on several geographic indicators and found significant trends. For these analyses, 'delivery_ratio' was the response variable. The 'delivery_ratio' variable captures how many of a customer's total orders are home delivery (e.g. if 3 of 10 orders are home delivery, 'delivery_ratio' is 0.3, or 30%). Using 'DistanceMiles' as the regressor variable (DistanceMiles represents the number of miles from the centroid of the customer's zip code to the location of the nearest Walmart store), this regression suggests that the correlation between a customer's proximity to a Walmart store and their home delivery adoption rate is positive and statistically significant. The above correlation explains 63.3% of the data, as per the adjusted R-squared of 0.633.

Other geographic considerations impacted home delivery behavior in other ways as well. The MIT team also looked at variables such as the number of Walmart customers per zip code, the density of Walmart customers per zip code (number of Walmart customers divided by total population).

Findings – Channel Choice Pipeline

The 801 survey responses were analyzed using a Random Effects Logit Model (RELM), where each customer–choice pair formed an observation. The MIT team focused on those that identified as having shopped via online grocery, which formed a group of 429 customers after outliers were removed.

Customer sensitivity to price combined with delivery window was quantified such that every increase in the level of delivery fee causes a person to be 20.7% less likely to choose home delivery.

Customer sensitivity to pick-up window was quantified such that this channel feature exhibited minimum moderating effect on channel choice.

Customer sensitivity to distance was quantified such that a customer is 2.77 times more likely to choose home delivery when a store is 15+ miles away as compared to a customer with a store that is less than 10 miles away.

Customer sensitivity to delivery agent was quantified such that delivery agent exhibited no moderating effect on channel choice.

Being a senior (65+ years old) reduces the likelihood that a customer is home delivery by 63.64% as compared to the youngest age group of 18 – 24.

Conclusion

Customer Profile Pipeline:

Location matters. There were statistically significant correlations between home delivery adoption and a customer's proximity to their nearest Walmart store, as well as density of Walmart stores and density of competitors. For example, the average customer living 10 miles from their nearest Walmart store orders home delivery approximately 3 times more frequently than the average customer 4 miles from their nearest Walmart.

The model based on the Customer Profile Pipeline produces a heat-map and corresponding ranked list of zip codes by the number of likely home delivery adopters. The ranked list provides a road-map for rolling out home delivery capabilities by detailing the critical markets for home delivery, and the heat-map allows Walmart the flexibility to adjust constraints and focus on regions that present particularly attractive opportunities.

Channel Choice Pipeline:

Price and distance matter. There were statistically significant correlations between home delivery channel choice and the cost of delivery combined with delivery window, as well as a customer's distance from their nearest Walmart. For example, while delivery agent is not a significant factor in channel choice, price is; every dollar increase in home delivery causes a person to be 20.7% less likely to choose home delivery.

Delivery agent does not matter. Whether a 3rd party such as Uber, Deliv or Walmart associates home deliver the order, the consumer choice won't be affected. The grocery retailer should seek the transportation service that delivers the required service level (quality of delivery, on-time delivery) at the minimum cost.

Pick-up from store window does not matter. Moving the window to the next day instead of same day won't affect the consumer's choice of channel. Hence, the grocery retailer can design its pick-up from store window based on minimum costs. With a pick-up from store window moved to the next day, the retailer can do overnight picking and avoid congestion of pickers and shoppers in the store.