Online Grocery & Omnichannel Strategy: Predicting Home Delivery Adoption

by

Ryan Alexander Alberts & Antoine Lahad Abinader

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 2018

Signed by Author……………………………………………………………………………………………
Ryan Alexander Alberts

Signed by Author……………………………………………………………………………………………
Antoine Lahad Abinader

Certified by…………………………………………………………………………………………………
Dr. Eva Ponce
Executive Director, MITx MicroMasters Program in Supply Chain Management
Capstone Project Supervisor

Accepted by…………………………………………………………………………………………………
Dr. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Online Grocery & Omnichannel Strategy: Predicting Home Delivery Adoption

by

Ryan Alexander Alberts & Antoine Lahad Abinader

Submitted to the Program in Supply Chain Management
on May 14, 2018 in Partial Fulfillment of the
Requirements for the Degree of Master of Engineering in
Supply Chain Management

ABSTRACT

The traditional way to reach customers in e-commerce is home delivery. Retailers have expanded fulfillment options to include picking up from a store, locker, 3rd-party collection point, and more. This study focuses on two channels: pick-up from the store and home delivery.

Groceries present a unique category for eCommerce due to particularly onerous complications from last-mile delivery of fresh products. Existing research is lacking in comparisons of channel options in the context of online grocery that capture interactions of channel and customer attributes. This study identifies critical markets for home delivery of online grocery and provides insights into drivers of channel choice in this context. It does so by first modelling home delivery adoption – applying machine-learning algorithms to historical customer data – and then analyzing channel preferences via a Discrete Choice Experiment devised by the authors expressly for this study.

The study quantifies the importance of geographic features in home delivery adoption, including density of existing online grocery customers and their distance from a store. The study also quantifies the likelihood of customer channel preference given varying channel attributes; for example, a customer is no more likely to choose pick-up from store if it is ready today vs. tomorrow.

Thesis Supervisor: Dr. Eva Ponce

ACKNOWLEDGEMENTS

The authors wish to express their gratitude and appreciation to Dr. Eva Ponce for charting the course of the research and for asking the questions, which brought out the important considerations relative to the subjects explored in this thesis.

The authors would also like to offer special thanks to Dr. Sina Golara whose expertise in choice modeling was essential to the success of this study.

The following individuals are thanked for their guidance and for providing their subject matter and industry expertise: Daniel Shea, Jody Ayers, Pushkar Raj, and Heather Gamper.

Finally, the authors thank the Supply Chain Management (SCM) Residential Class of 2018 for making their time in Cambridge a memorable one:

Hanin Aboutaleb
Ali Al-Habib
Nina Arizaleta Valera
Sun Assavaniwej
Xiwen Bai
Anushka Budhiraj
Xime Castañon Choque
David Cheung
Jeff Du
Nicolas Favier Gonzalez
Geri Felicio
Mary Kate Gorman
Harshvardhan
Nacho Helguera
Mita Hendra Mukti
Abou Rami
Evan Humphrey
Aline Ingberman
Puneet Jain
Federico Laiño
Alan Li
Tony Li
Yiyao Li
Ben Lillijord
Melissa Medina
Russ Miller
Bernadette Orende
William Phillips
Ross Pieper
Christos Rallis
Deepika Sharma
Lauren Sittler
Eugene Sohn
Hilary Taylor
Ian Teoh
Anshu Thakur
Ben Yin
Darryl Yau
Justin Yoon
Xiaohan Zhu
Krista Hoffman
Dave Rohdes
Jeff Bilodeau

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. 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" ("OmniChannel" 2017). The abundance of smartphones, growth of internet penetration, and innovation in technology are driving omnichannel retailing, where shoppers spend 15-30% more than traditional retail (Yee, Dahiya, & Kraemer, 2015).

For Walmart, groceries are defined as fresh, perishable items such as produce, meat, dairy, etc. as well as pantry items such as canned food, consumables, pet food, and the like. Groceries present a unique category for eCommerce due to entrenched customer preferences that have been shaped by generations of routine, solidified by way of frequent, habitual trips to the grocery store. They are also particularly onerous for home delivery – the benchmark of ecommerce – due to complications from last-mile delivery of fresh products.

Walmart currently has the pick-up in store channel in 1,128 stores, 38 of which are equipped with home delivery capabilities as of early 2018. 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. Figure 1 represents different distribution channel options for online orders.

The traditional way to reach customers in e-commerce is home deliveries (represented by the blue arrows in Figure 1). More recently, retailers have started to offer different options for fulfilling online orders, including picking up from the retail store (red flow in Figure 1), picking up from lockers (automated package stations, APS, green flow in Figure 1), and picking up from collection points (e.g. convenience stores, gas stations, etc.) (yellow flow in Figure 1). This capstone project focuses on two channels: pick-up from the store and home delivery.

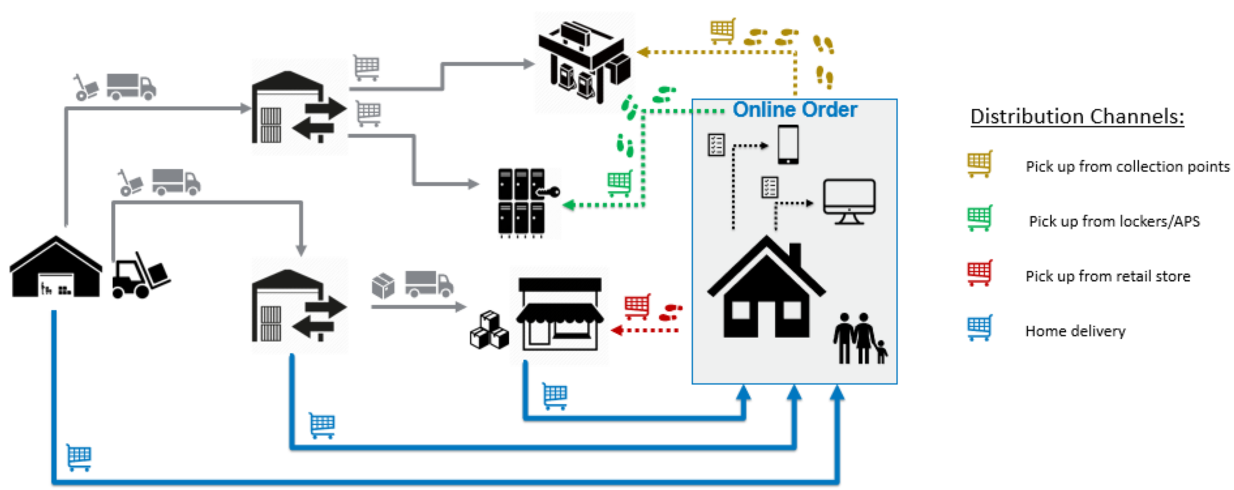


Figure 1-1: Distribution Channels for Online Orders

Walmart operates two distinct online platforms for grocery sales: Walmart.com and grocery.walmart.com. The former offers general merchandise as well as non-perishable grocery items such as dry pantry goods like cereal and snacks, while the latter focuses on perishables, such as dairy and frozen goods. Any Walmart customer can shop online and ship non-perishables to their home, but only a limited subset have access to grocery home delivery.

MIT's research team has partnered with Walmart's Global Customer Insights & Analytics (GCIA) & Online Grocery Strategy and Development (OGSD) teams to explore omnichannel strategy for these two channels (pick up from store and home delivery). Currently, fulfillment of online grocery orders by home delivery is offered in limited locations, and launching the 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 Walmart's 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?

The first objective is achieved through the Customer Profile Pipeline (sections 3.1 and 4.1), via a model that produces predictions regarding geographic areas with the highest likelihood of home delivery channel adoption. Using existing customer purchasing behavior data, the MIT team designed an algorithm based on the regions where both channels, home delivery and pick-up from store, are active. The model built was then deployed to predict the home delivery adopters using the data of the customers located in regions with pick-up from store channel only. Model results are explored in sections 4.1.4 and 4.1.5. The results included a map depicting expected home delivery adopters associated with each zip code in the data set. Section 4.1.6 illustrates two heat maps where the predictions were aggregated by state, and Appendix D lists the top 20 zip codes by volume and density of home delivery adopters.

The second objective is achieved through Channel Choice Pipeline (section 3.2 and 4.2), via a Discrete Choice Experiment (DCE) that evaluates the effect of channel features and demographics on customer channel choice. The MIT team devised a DCE consisting of a survey of the country's general population. The methodology is explained in section 3.2. The survey 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. The results are explored in section 4.2.1.

The MIT team's study identifies the "hot" markets for online grocery shopping with home delivery service and provides insights into drivers of channel choice capable of guiding the grocery retailer towards channel demand shaping.

2. LITERATURE REVIEW

The first part of this chapter provides background and reviews the existing literature surrounding drivers of channel choice for online grocery. So far, there is limited research comparing customers preferences between home delivery and in store pick-up channels.

The second part of this chapter explains the Discrete Choice Experiment (DCE). The MIT team reviewed Verma, Plaschka, Hanlon, Livingston, & Kalcher (2008), McFadden (1973) Quaife (2016) who explain the method's advantages and limitations. Recommendations on the impact of sample size on Discrete Choice Experiments were taken into consideration from Vilikus (2014), McCullough (2002) and Orme (2010).

2.1 Online Grocery and Channel Choice

Pradeep Chintagunta et al. (2012) found that (1) travel cost discourage customers from shopping instore (the farthest they are from their favorite physical store, the more likely they will shop online), (2) when customers have many items to buy or the items are heavy, the efficiency gain from online shopping is considerable. (3) Delivery charges have a strong discouraging effect on customers: although their trip to the store has a cost, they are reluctant to pay a delivery fee for home delivery and (4) customers are likely to visit the store to purchase perishable items.

Chintagunta et al. (2012) show that cost sensitivity varies with demographics:

- Larger families have a stronger preference for shopping online
- Larger families are more sensitive to delivery charges

The age factor is reflected in Nielsen Global E-commerce and New Retail Report (2015), which surveyed roughly 30,000 customers in 60 countries to determine online shopping habits across age groups. Their results are illustrated in Appendix B: Online Grocery Nielsen Survey (2014).

Droogenbroeck et al. (2017) explored the personal and household characteristics that motivate customers towards the pick-up in store channel. Characteristics such as high educational level, presence of young children and number of full-time employed adults within the household drive customers to this channel.

Gao & Su (2016) studied the impact of buying online and pick-up from store on store operations. This channel helps consumer choice by providing real-time information about product availability and by reducing the hassle cost of shopping. In their conclusion, Gao & Su (2016) specify that it may not be profitable to implement pick-up from store channel on products that are selling well in store. In addition, customers initiating an online order and finding the desired item is out of stock will not visit the store. However, having an additional channel will help retailers expand their market coverage, however, pick-up from store may cannibalize sales from other more profitable chains. Grocery retailers should be aware of the revenue and cost implications of new distribution channels and after understanding the drivers behind channel choice, retailers should seek to shape each channel demand in order to maximize profits.

An interesting research conducted by Rabinovich, Sousa, Park, & Golara (2018) studies the effect of removing the delivery fee of online orders picked up from the store. In the study, customers are segmented into two groups, group 1 includes customers who use exclusively pick-up from store channel and group 2 includes customers who use both home delivery and pick-up from store channels. The study concludes that only group 1 weekly spend increases after removing the fee. There was no evidence to suggest that the fee elimination led to an increase in the weekly spend of group 2 via pick-up from store channel. In addition, both groups increased the frequency of their orders and decreased their order size. Interestingly, Rabinovich, Sousa, Park, & Golara (2018) found that the increase in revenue from group 1 customers is not large enough (1) to cover the operating costs resulting from fulfilling a higher volume of orders and (2) to make up for the revenue from the pick-up from store fees no longer being collected by the retailer. Currently, Walmart does not charge any fee for the pick up from store channel and the MIT team did not tackle this channel feature in this research.

In her paper "A Consumer Perspective on Grocery Retailers' Differentiation", Bellini (2015) explored consumers' perception of grocery retailers in Italy to understand what shopping needs consumers want to satisfy when they choose a grocery store and which store attributes consumers perceive different between retail store formats and between grocery retailers. As a result, the needs of grocery shoppers are time saving, money saving, trust and quality of the shopping experience. Bellini (2015) argues that these needs can be satisfied by more than one store format. For example, consumers can save time when shopping in small convenience store a well as hypermarkets which have developed in-store services such as fast check-outs. Bellini (2015) then concludes that range and price are no longer enough to meet consumers' shopping needs, retailers should manage other retail levers: in-store marketing, services and technology. This conclusion confirms the importance for retail store to diversify their sales and distribution channels and leverage technology to provide a more seamless shopping experience, both online and offline.

2.2 Choice Modeling

Verma, Plaschka, Hanlon, Livingston, & Kalcher (2008) paper presents an overview of using discrete choice modeling for service sector application. Discrete choice modeling is increasingly being used in many applications for the service sector to predict customer choice. The authors clarify that economic choice theory assumes that individuals' choice behavior is generated by maximization of preferences or utility. Utility is defined as ‘‘judgments, impressions, or evaluations that decision makers form of products or services, taking all the determinant attribute information into account".

McFadden (1973) specifies the four stages of a Discrete Choice Experiment (DCE) design. The four stages are (1) Identification of attributes and levels (2) Experimental design (3) Data collection and (4) Data analysis. While there are many so-called “Stated Preference” methodologies – conjoint analysis, contingent valuation, etc. – DCEs are effective in consumer retail space due to the existence of categorical dependent variables, i.e. discrete choices, which can be modelled by logistic regression and conditional logit analysis. The DCE is subject to several limitations. Hypothetical bias is an issue, because the DCE collects stated rather than revealed preferences; however, this bias has been addressed and overcome by a number of methods governing survey architecture. There is also the issue of omitted variable bias, where some attribute driving a choice is not stated in the choice sets but is nevertheless taken into consideration by the respondent.

Quaife (2016) further explains some limitations to Discrete Choice Experiments (DCEs). DCEs set up hypothetical choices which may induce hypothetical bias. Also, using too many attributes will complicate the DCE design and increase the risk of participants’ fatigue.

Sample size determination is important to reduce sampling and measurement errors. In their papers, Vilikus (2014), McCullough (2002) and Orme (2010) highlight that there are no formula that give an accurate estimation of the needed sample size to fulfill the goals of a DCE with a high degree of confidence. Their recommendations are based on rules of thumb and their experiences in previous conjoint analysis studies. McCullough (2002) states that a sample size as low as 75 respondents can create a reliable model. However, he clarifies that 75 is the minimum number to examine one analytic cell: separating male and female respondents requires 75 male respondents and 75 female respondents. Orme (2010) reports a rule of thumb developed by Johnson (author of Sawtooth Software's Choice Based Conjoint System) to determine minimum sample size for aggregate level full-profile Choice Based Conjoint (CBC) modeling:

$$\frac{nta}{c} \geq 500$$

where n is the number of respondents, t is the number of tasks, a is number of alternatives per task, and c is the number of analysis cells. Orme (2010) concludes that sample sizes for conjoint studies generally range from about 150 to 1,200 respondents and recommends at least 300 respondents for robust quantitative research, and between 30 – 60 for analyzing subgroups.

2.3 Groundwork for Current Research

The literature review provides a nuanced perspective of channel choice as a function of attributes of the customer. However, attributes of the channel (delivery price, window, etc.) can be as significant; specifically, those attributes that Walmart seeks to leverage to gain market share are underrepresented in existing literature. Existing literature is lacking in comparisons of channel options in the context of online grocery capture interactions of channel and customer attributes to satisfy Walmart's strategy.

3. 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. This data serves as the input to an algorithm that predicts home delivery adoption, defined in section 3.1.2.2.

The second is the channel choice pipeline, flowing from a Discrete Choice Experiment (DCE) devised by the MIT team, consisting of a survey of Walmart 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.

3.1 Customer Profile Pipeline

3.1.1 Data Wrangling

The team analyzed historical records accessed by Walmart's GCIA team from company servers, which capture critical purchasing behavior data. The types and structures of data sources available therefore governed the range of feasible methods.

Purchasing behavior and order history was then aggregated by customer for

2,253,976 unique Walmart customers in the USA who ordered at least once between the first week of 2017 and the first week of March 2018. The team supplemented the historical sales data provided by Walmart with geographic, census, and engineered data described in detail in the Applied Definitions Section below.

3.1.2 Applied Definitions

3.1.2.1 Historical Data Features

Each record in the historical data set represents one unique customer who placed at least one Online Grocery order via Walmart.com platform. Online groceries orders on Walmart.com can be delivered via two channels only: home delivery or pick-up from Walmart store. This data set consists of features defined in Table A-1, Table A-2 and Table A-3 of Appendix A. In addition, The MIT team engineered features listed in Table 3-1.

Table 3-1: Engineered Features for the Historical Data

Feature	Description
zip_customer_count	Number of Walmart Online Grocery customers per zip code
zip_mean_sales	Average of *mean_sales* aggregated by zip code. All customers are weighted equally (USD per order)
delivery_ratio	*(del_ord_cnt) / (del_ord_cnt + pickup_ord_count)*
pop_density	*population / land area* in inhabitant per square miles
pop_density_round	*pop density* rounded to the nearest thousand
mean_sales_delta	*mean_sales / zip mean sales*
mean_item_value	*mean_sales / avg_ord_size* (USD per item)
comp_density	*total comp / population (*number of competitors per inhabitants)

distance_value	*mean_sales / distance miles* (USD per mile)
walmart_density	*zip customer count / population*
delivers_binary	0 if the store has the pick-up from store channel only 1 if the store has the pick-up and home delivery channels
GO_Date	Date of launch of online grocery in the store

3.1.2.2 Target Variable: Home Delivery Adopters

The MIT team segmented Walmart.com online grocery customers by their total number of home deliveries and orders picked up in store, represented in Table 3-2. The customers in red ordered less than once from one or both channels and are designated "non-adopters" of online grocery. The customers in yellow ordered more than twice from the pick-up from store channel and are defined as pick-up adopters. The customers in green ordered more than twice via home delivery and are defined as home delivery adopters. The target variable is home delivery adopters (customers in green), whose behavior and preferences are the subject of the current research. Customers in red are excluded from the model.

Table 3-2: Walmart.com Online Grocery Customers' Segmentation

		Pick-up Order Count		
		0	1	2+
Home Delivery Order Count	0	Not Online Grocery customer	Trial Pick-up	Repeat Pick-up Only
	1	Trial Delivery	Uncommitted customer	Skew Pick-up
	2+	Repeat Delivery Only	Skew Delivery	Repeat Online Grocery customer

3.1.3 Cluster Analysis

The objective of the cluster analysis was to explore features relevant to home delivery adoption. The MIT team ran k-means algorithm for clustering on a selection of features from historical data set, aggregated by zip code. K-means procedure splits the data into k clusters. Each cluster has a centroid that corresponds to the mean value for the members in that cluster. The objective of the algorithm is to minimize the total of the sum of distances between cluster centroids and members of such clusters.

Lloyd's algorithm was used with squared Euclidean distances to compute the k-means clustering for each k (Kanungo, 2002). Combined with the splitting procedure to determine the initial centers for each $k > 1$, the resulting clustering is deterministic, with the result dependent only on the number of clusters. Different values of k were tested to suggest an optimal number of clusters using the Calinski-Harabasz criterion to assess cluster quality. To evaluate and compare models with different feature selections, the MIT team investigated three metrics:

1. Between-group sum of squares: a metric quantifying the separation between clusters as a sum of squared distances between each cluster's center (average value), weighted by the number of data points assigned to the cluster, and the center of the data set. The larger the value, the better the separation between clusters.

2. Within-group sum of squares: a metric quantifying the cohesion of clusters as a sum of squared distances between the center of each cluster and the individual marks in the cluster. The smaller the value, the more cohesive the clusters.
3. Total sum of squares: totals the between-group sum of squares and the within-group sum of squares.

The MIT team selected the model with the maximum ratio between between-group sum of squares over total sum of squares. This ratio gives the proportion of variance explained by the model.

3.1.4 Classification Models

In order to produce predictions regarding geographic areas with the highest likelihood of home delivery channel adoption, the MIT team relied on open-source, supervised machine-learning algorithms. Specifically, the team relied on classifiers, a class of algorithm that identifies to which subset of categories a new observation belongs, using a training set of data whose membership is known. The algorithm learns using explanatory variables, or features, that can be continuous, categorical, ordinal, etc.

The algorithms used include the following python libraries: LightGBM, sklearn (KNN and logistic regression), XGBoost, and statsmodels. The target variable was a binary class that identified whether a particular observation was a customer defined as a home delivery adopter. The model used a set of 18 features including those engineered by the

MIT team. The model produced a prediction representing the likelihood of a particular observation (customer profile) to be among the class of customers that in fact are home delivery adopters.

After being trained, the model then makes predictions about general populations that it has not previously encountered. The predictions were used to pinpoint geographic regions that were most likely to contain positively identified customers, i.e. those that would most likely engage in home-delivery of online grocery. With these geographic locations identified using Walmart's existing online grocery customers, the MIT team created a heat-map that visualizes which zip codes should be equipped with home delivery capabilities, according to which store fall within geographic zones previously identified as highly dense in home delivery adopters. The model ultimately uses these predictions, aggregated by geographic area, to enable a visual representation indicating critical markets for launching home delivery capabilities.

3.2 Channel Choice Pipeline

Discrete Choice Experiments (DCE) are a means of making statistical inferences about a population's choice behavior, by relating the choices made with the attributes of mutually exclusive alternatives. Intuitively, when you only observe the outcome of a choice, without explicitly attributing it to some specific cause, the testable implications of the choice are obscured (McFadden, 1973). However, the DCE approach is structured to explicitly state the objects of choice – i.e. the attributes of each alternative, the attributes of the decision-makers, and the actual choice.

3.2.1 Discrete Choice Experiment Design

Walmart contracted Harris Poll[1] via Nielsen Media[2] to conduct a survey of the US general population of online grocery customers. The sample set was meant to be representative of a national scale of both rural and urban customers. Our key sample significance test consisted of ensuring a minimum of 800 respondents.

The survey consists of introductory demographic questions followed by choice sets and is included as Appendix F. The introductory questions were used to classify the respondents' familiarity and willingness to engage online grocery:

1. Customers that had shopped online grocery before,
2. Those that had never shopped online grocery, but are interested in doing so, and
3. Those that had never shopped online grocery but are not interested in doing so[3].

Choice sets are presented in random sequence of monadic cards, with each choice set presenting the same delivery options and the same attributes, except with varying attribute values. A matrix of attributes and attribute values can be seen in Table 3-3:

[1] https://theharrispoll.com
[2] http://www.nielsen.com/us/en.html
[3] This group was prompted to exit the survey and was not included in the experiment.

Table 3-3: Attributes and Levels for Choice Sets

Attribute	Level		
Home Delivery/Pick-up from Store window	Order placed by 1pm and delivered/pick-up as soon as one hour	Order placed by 1pm and delivered/pick-up as soon as 4 hours	Order placed anytime today and delivered/pick-up the next day
Home Delivery agent	3rd Party (e.g. Uber/Drive)	Walmart associates	
Home Delivery Cost	\$6.99	\$9.99	\$14.99
Store distance to home	less than 10 miles	10 to 15 miles	more than 15 miles
Store on commute	Yes	No	

4. FINDINGS

4.1 Customer Profile Pipeline

The first subsection describes the current grocery home delivery markets of Walmart and the segmentation of the data set into training and testing sets. Then, a cluster analysis is conducted on Colorado which is a mature market for both channels, home delivery and pick-up from store. The cluster analysis reveals the combined effect of population density, competitor density and mean sales. It allows the exploration of feature sets that would be useful for the predictive model. The following subsections explain classification models that were tested (subsection 4.1.3) and the model selection (subsection 4.1.4). Finally, subsection 4.1.5 reveals the results of the predictive model: tables and heat maps of the total predicted number of home delivery adopters aggregated by zip code and state.

4.1.1 Data Segmentation

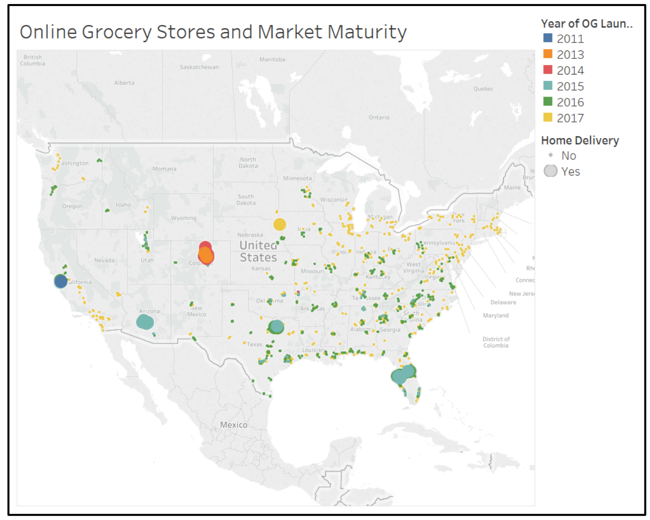


Figure 4-2: Online Grocery Store & Market Maturity

Walmart has 1,127 stores with online grocery capabilities providing the pick-up from store channel. 38 out of the 1,127 stores also provide the home delivery channel. Figure 2 shows the location of these stores and the year when Online Grocery service was launched. The MIT team began by segmenting the historical customer data by channel access. The resulting two segments consisted of 5 discrete geographic areas of totaling 61,494 customers that had access to both channels and 1,629,086 customers that had access to only the pick-up channel.

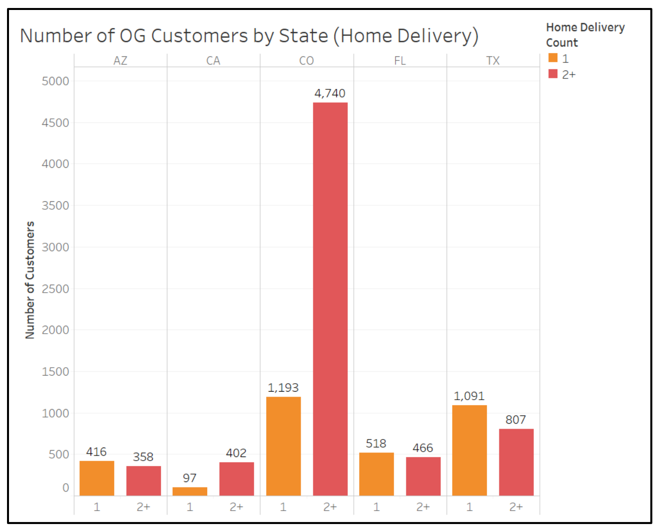


Figure 4-1: Online Grocery Customers Grouped by State and Home Delivery Count

The bar graph in Figure 3 shows the number of customers who have access to both channels and it is segmented by state (AZ, CA, CO, FL, TX) and home delivery count (1 in orange and 2+ in red). As defined in section 3.1.2.2, a customer with a home delivery count of 2 or more reflects a home delivery adopter. Of the 61,494 customers that had access to both channels, 42,651 had two or more online orders and served as the training set for the predictive models. All models then split the data 80:20, where 80%, or 31,988 customers were used for the training set and 20%, or 10,663 customers were used as the test set. Of those 10,663 test customers, 8,996 were not home delivery adopters and 1,667 were. The test set was used to score the predictive capacity of models.

4.1.2 Cluster Analysis

The state of Colorado has the highest number of home delivery adopters (Figure 3), with an online grocery service in 13 stores since 2013-2014. Therefore, the MIT team focused on Denver, CO, which is currently the largest and most mature online grocery home delivery market for Walmart. The MIT team ran a k-means cluster analysis on population density per zip code ('pop_density'), number of online

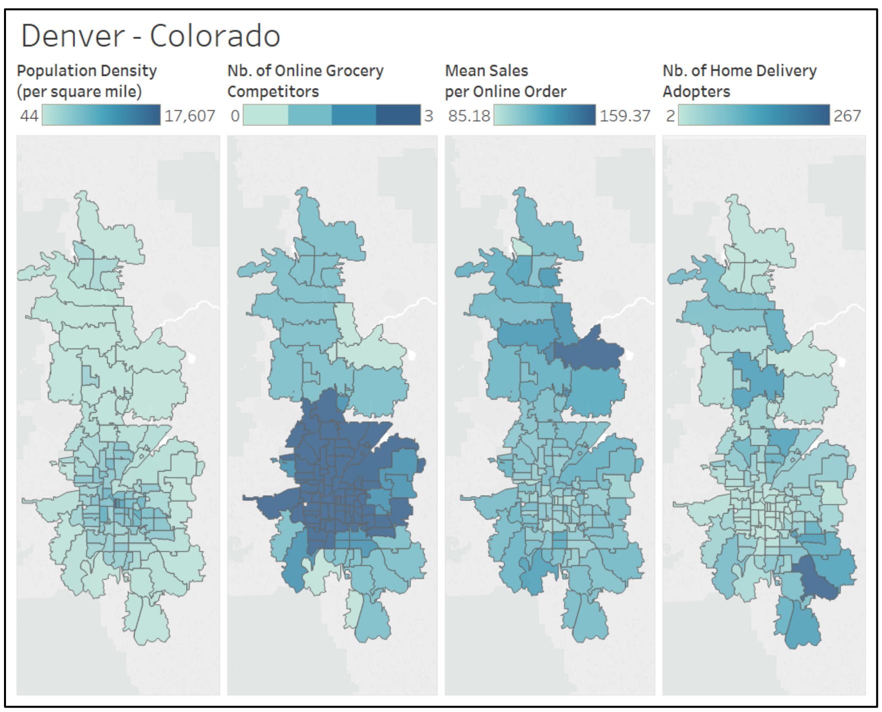


Figure 4-3: Relevant Features for K-Means

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 (Figure 4).

The k-means algorithm outputs two clusters with a maximum ratio of (between-group sum of squares)/(total sum of squares) = 0.569 (Table 4-1).

Table 4-1: Summary Diagnostics

Number of Clusters:	**2**
Number of Points:	107
Between-group Sum of Squares:	10.63
Within-group Sum of Squares:	8.0478
Total Sum of Squares:	18.677

The blue cluster 1 (Figure 5) groups the city area, defined by high population density and number of competitors, as well as low mean sales per customer. Meanwhile, cluster 2, in orange, covers suburbs where there is low population density, low number of competitors and high mean sales per customer. Table 4-2 reveals that home delivery adopters are more likely in cluster 2 in orange than cluster 1.

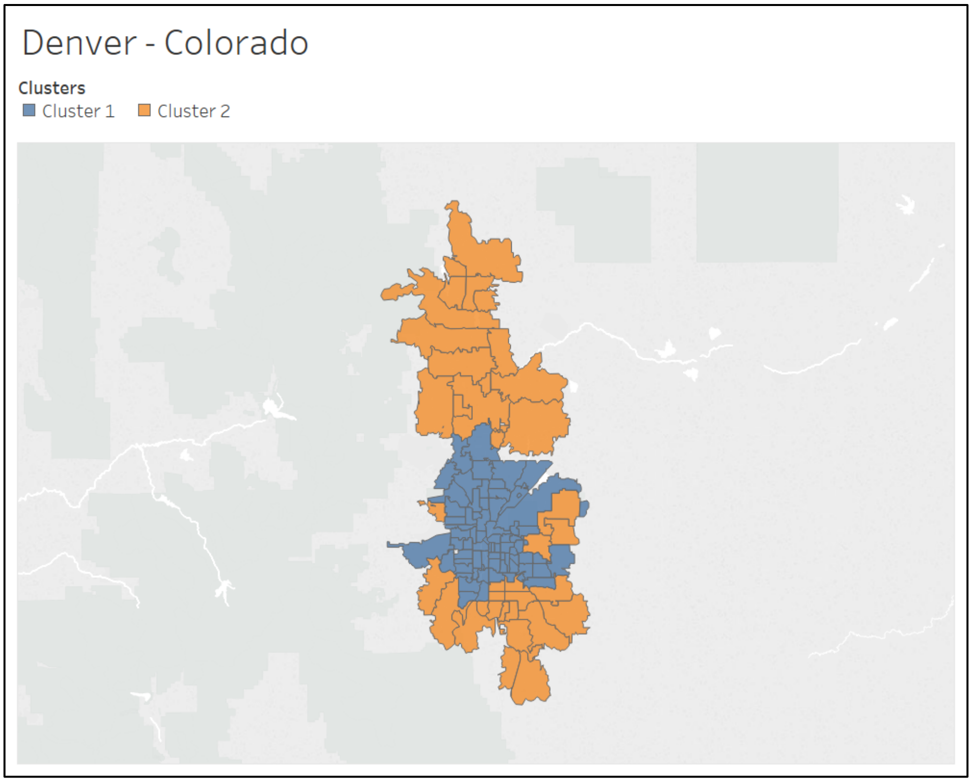


Figure 4-4: Clusters from K-Means Analysis

Table 4-2: Summary Statistics

		Centers			
Clusters	**Number of Items**	**Avg. Pop Density**	**Avg. Mean Sales**	**Avg. Total Comp**	**Sum of Number of Records**
Cluster 1	69	4422.3	122.61	2.9855	34.623
Cluster 2	38	1249.7	137.99	1.1316	61.684

4.1.3 Classification Models Overview

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.

The first was a gradient boosting machine (GBM) with a binary classification objective function, using weak-learning decision trees iteratively for 100 rounds according to binary logarithmic loss. Log loss quantifies the uncertainty of a prediction by penalizing false classifications. Minimizing the Log Loss is roughly equivalent to maximizing the accuracy of the classifier. The function is defined as

$$d(x, y) = \sum_{i=0}^{N-1} |x_i - y_i|^{1/p}$$

Where:
N = Number of sample instances
M = number of possible labels
y_{ij} = binary indicator identifying correct classification for instance i
p_{ij} = model probability of assigning label j to instance i

The second was k-Nearest Neighbors, which is a non-parametric majority vote of closest observations according to Euclidean distance. This model has no explicit training phase and no feature distribution assumptions. The distance function that determined class membership is defined as:

$$-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij} \log p_{ij}$$

Where:
N = Number of Observations
x_i = Observation x at point i
y_i = Observation y at point i

The third is a Naïve Bayes model, which is a probabilistic classifier that applies Bayes Rule:

$$P(y \mid x_1, \dots, x_n) \propto P(y) \prod_{i=1}^{n} P(x_i \mid y)$$

$$\Downarrow$$

$$\hat{y} = \arg\max_{y} P(y) \prod_{i=1}^{n} P(x_i \mid y),$$

Where:

y = Class Variable

x_i = Feature vector x at point i

$\hat{y}$ = Predicted y at point i

4.1.4 Model Selection

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.

Gradient Boosting Machines		Predictions		
		Precision	Recall	Support
Actuals	0	0.92 *(8,428)*	0.94 *(568)*	*(8,996)*
	1	0.63 *(689)*	0.59 *(978)*	*(1,667)*
		f1 score = 0.88		*(10,663)*

The GBM model stood out in accurately predicting 58.66% of home-delivery adopters – 978 out of 1,667 – as compared to the kNN and Naïve Bayes classifiers that correctly predicted 45.11% and 44.39%, respectively. The GBM model was superior in terms of the above measure, Recall, in addition to the other component of the F1 score, Precision: the GBM model scored 92.44% in accurately predicting non-home-delivery adopters – 8,428 out of 9,117 – as compared to the kNN and Naïve Bayes classifiers that correctly predicted 90.36% and 90.31%, respectively (Appendix C: Confusion Matrices)

This 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).

4.1.5 Feature Ranking

The GBM model produced a number of meaningful results in the form of feature importance (**Figure 6** and **Figure 7**).

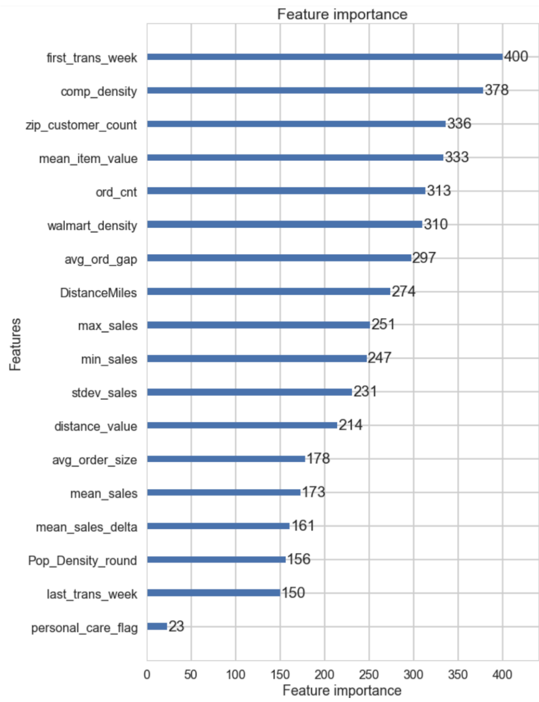


Figure 4-5: GBM – Feature Importance

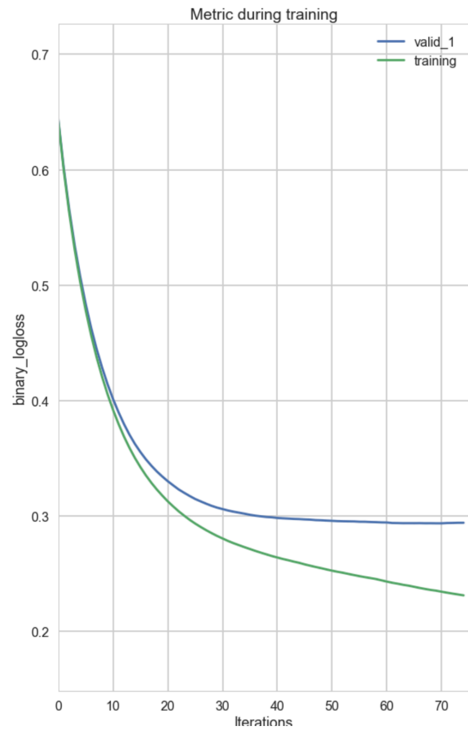


Figure 4-6: Evaluation of GBM Model – Binary Log Loss

Feature importance refers to the relative weight of each feature in the model, i.e. predicative capacity. Each feature is ranked according to how many times it was used to correctly classify previously misclassified members of a class, in terms of the number of decision tree nodes it factored into when classifying the test set. The features can then be compared for their relative importance, in the context of the overall quality of the model. Looking at the top features by feature importance, the MIT team found significant trends.

First: Walmart location matters. 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, as OLS regression requires a continuous 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 'delivery_ratio' as the response variable and '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 captures the relationship between a customer's proximity to a Walmart store and their home delivery adoption rate (in the above example, the adoption rate is 30%). The MIT team bucketed 'DistanceMiles' by mile, so that the regression considered mean 'delivery_ratio' per mile. For example, the average 'delivery_ratio' for customers whose nearest store was 10 miles was approximately 0.3, or 30%.

The graph in Figure 8 plots the expected 'delivery_ratio' for customers that fall within a given distance from their nearest Walmart (e.g. a customer that lives 3 miles from the nearest Walmart has an expected delivery adoption rate of approximately 0.10 or 10%). The above correlation explains 63.3% of the data, as per the adjusted R-squared of 0.633.

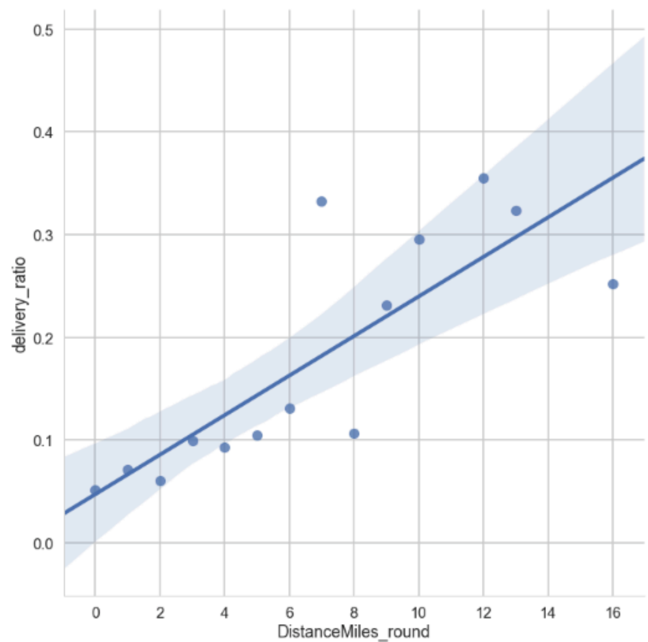


Figure 4-7: Delivery Ratio as a Function of Distance Miles

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          delivery_ratio   R-squared:                       0.661
Model:                             OLS   Adj. R-squared:                  0.633
Method:                  Least Squares   F-statistic:                     23.40
Date:                 Thu, 03 May 2018   Prob (F-statistic):           0.000407
Time:                         16:46:17   Log-Likelihood:                 18.760
No. Observations:                   14   AIC:                            -33.52
Df Residuals:                       12   BIC:                            -32.24
Df Model:                            1
Covariance Type:             nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
Intercept                 0.0472      0.033      1.436      0.177      -0.024       0.119
DistanceMiles_round       0.0192      0.004      4.837      0.000       0.011       0.028
==============================================================================
Omnibus:                         2.333   Durbin-Watson:                   2.282
Prob(Omnibus):                   0.312   Jarque-Bera (JB):                0.797
Skew:                            0.557   Prob(JB):                        0.671
Kurtosis:                        3.353   Cond. No.                         15.0
==============================================================================
```

Figure 4-8: Summary Statistics for Figure 8

With a P-value of less than 0.000, the likelihood of a customer's 'delivery_ratio' falling outside the expected range (variability denoted by shading, above) is extremely low: P>|t| = 0.000. With a Prob(F-statistic) = 0.0004, the probability that the model fits the data is high. The MIT team interpret this result as suggesting that there is a strong correlation between a customer's proximity to a Walmart store and their home delivery adoption rate (Figure 9).

Geographic considerations impacted home delivery behavior in other ways as well. The MIT team followed a similar process performing OLS regression with 'delivery_ratio' as the response variable using several explanatory variables, including 'zip_customer_count' and 'walmart_density'. 'zip_customer_count' captures the total number of Walmart customers in any given zip-code. For 'zip_customer_count', the MIT team again bucketed the explanatory variable so that the regression considered mean 'delivery_ratio' per 'zip_customer_count' group (in this case, groups of 10, i.e. zip codes with between 5 – 15 Walmart customers, 15 – 25 Walmart customers, and so on). The team found that more Walmart customers means higher home delivery adoption:

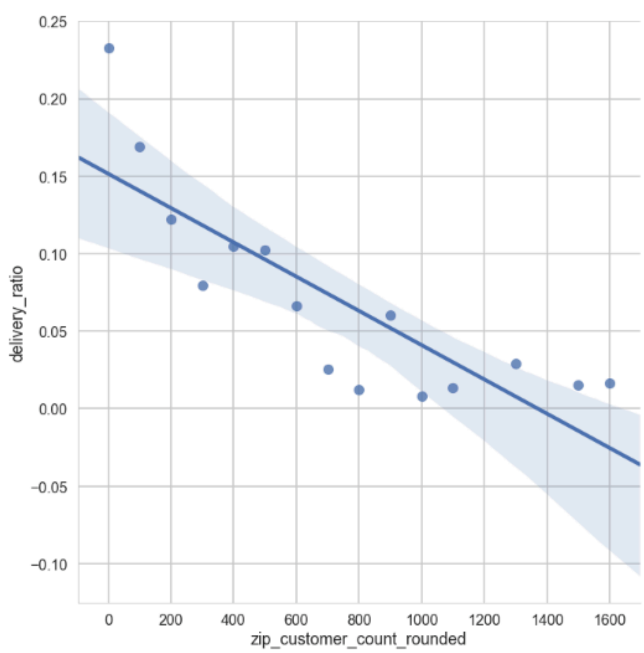


Figure 4-9: Delivery Ratio as a Function of Walmart Customer Count

Figure 10 plots the expected ‘delivery_ratio’ for customers that live in zip codes with a given number of Walmart customers (e.g. a customer that lives in a zip code with 400 Walmart customers has an expected delivery adoption rate of approximately 0.11 or 11%). The above correlation explains 67.0% of the data, as per the adjusted R-squared of 0.670.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:         delivery_ratio   R-squared:                       0.694
Model:                            OLS   Adj. R-squared:                  0.670
Method:                 Least Squares   F-statistic:                     29.44
Date:                Thu, 03 May 2018   Prob (F-statistic):           0.000116
Time:                        23:07:13   Log-Likelihood:                 28.816
No. Observations:                  15   AIC:                            -53.63
Df Residuals:                      13   BIC:                            -52.22
Df Model:                           1
Covariance Type:            nonrobust
=============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------
Intercept                     0.1515      0.018      8.474      0.000       0.113       0.190
zip_customer_count_rounded   -0.0001   2.04e-05     -5.426      0.000      -0.000   -6.65e-05
==============================================================================
Omnibus:                        1.032   Durbin-Watson:                   0.802
Prob(Omnibus):                  0.597   Jarque-Bera (JB):                0.607
Skew:                           0.474   Prob(JB):                        0.738
Kurtosis:                       2.730   Cond. No.                     1.60e+03
==============================================================================
```

Figure 4-10: Summary Statistics for Figure 10

With a P-value of less than 0.000, the likelihood of a customer’s ‘delivery_ratio’ falling outside the expected range (variability denoted by shading, above) is extremely low: P>|t| = 0.000. With a Prob(F-statistic) = 0.0001, the probability that the model fits the data is high. The MIT team interpret this result as suggesting that there is a strong correlation between the number of Walmart customers in a particular zip code and the home delivery adoption rate of a customer in that zip code.

Another significant geographic consideration is ‘walmart_density’. ‘walmart_density’ captures Walmart customers as a percentage of the total population. For ‘walmart_density’, the MIT team again bucketed the explanatory variable so that the regression considered mean ‘delivery_ratio’ per ‘walmart_density’ group (in this case, zip codes with densities of 0.001 – 0.002, 0.002 – 0.003, and so on). The team found that more Walmart customers means higher home delivery adoption:

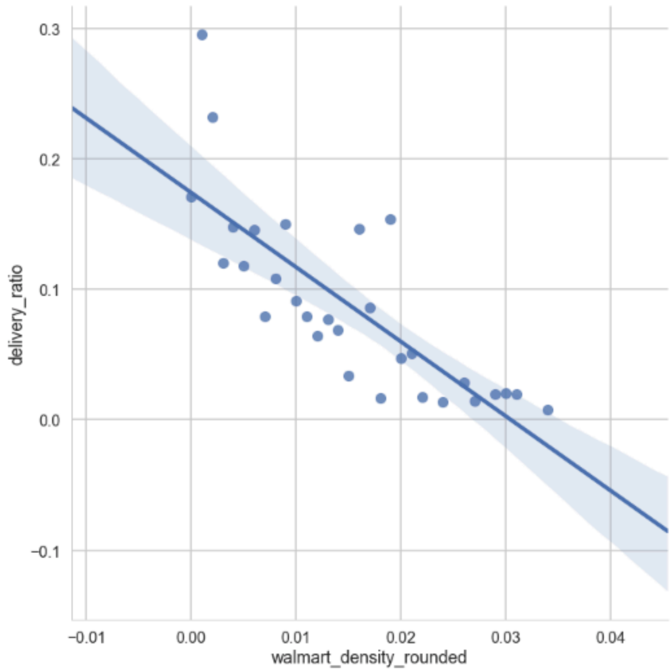


Figure 4-11: Delivery Ratio as a Function of Walmart Customer Density

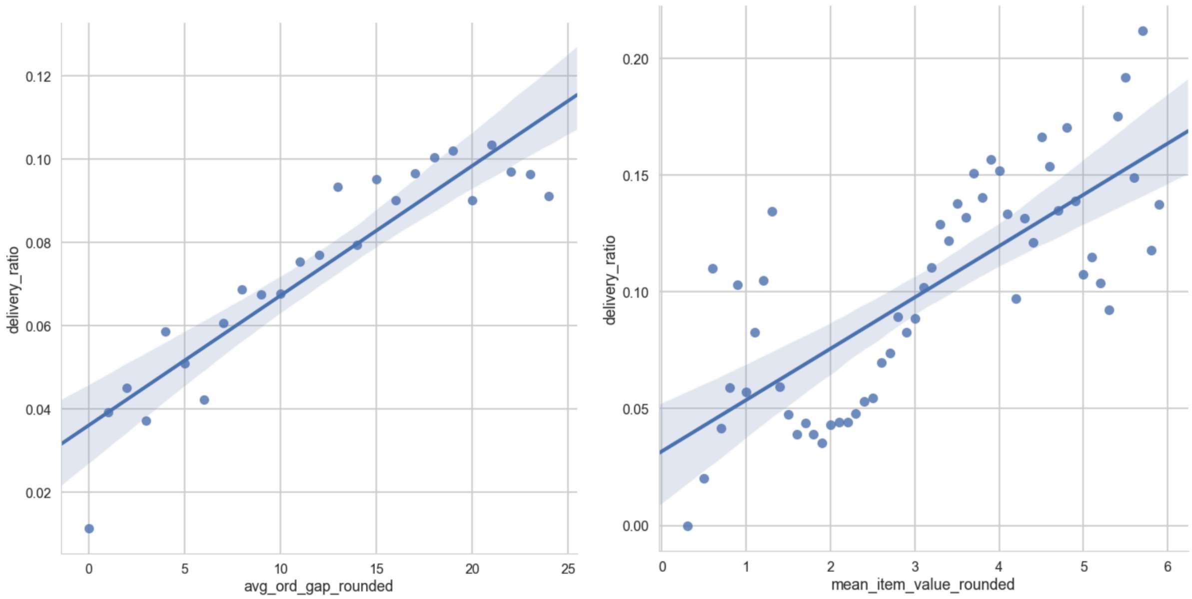
Figure 12 plots the expected ‘delivery_ratio’ for customers that live in zip codes with a given density of Walmart customers (e.g. a customer that lives in a zip code with 1 Walmart customer per 100 residents has an expected delivery adoption rate of approximately 0.09 or 9%). The above correlation explains 61.6% of the data, as per the adjusted R-squared of 0.616.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          delivery_ratio   R-squared:                       0.630
Model:                             OLS   Adj. R-squared:                  0.616
Method:                  Least Squares   F-statistic:                     47.59
Date:                 Thu, 03 May 2018   Prob (F-statistic):           1.69e-07
Time:                         17:03:51   Log-Likelihood:                 52.505
No. Observations:                   30   AIC:                            -101.0
Df Residuals:                       28   BIC:                            -98.21
Df Model:                            1
Covariance Type:             nonrobust
===========================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------
Intercept                   0.1740      0.015     11.732      0.000       0.144       0.204
walmart_density_rounded    -5.7109      0.828     -6.899      0.000      -7.407      -4.015
==============================================================================
Omnibus:                        10.294   Durbin-Watson:                   1.869
Prob(Omnibus):                   0.006   Jarque-Bera (JB):                8.893
Skew:                            1.182   Prob(JB):                       0.0117
Kurtosis:                        4.235   Cond. No.                         104.
==============================================================================
```

Figure 4-12: Summary Statistics for Figure 12

With a P-value of less than 0.000, the likelihood of a customer's 'delivery_ratio' falling outside the expected range (variability denoted by shading, above) is extremely low: P>|t| = 0.000. With a Prob(F-statistic) = 0.0000002, the probability that the model fits the data is high. The MIT team interpret this result as suggesting that there is a strong correlation between the density of Walmart customers in a particular zip code and the home delivery adoption rate of a customer in that zip.

In addition to geographic indicators of home delivery, the team was able to establish that customers who use home delivery tend to order less frequently, to stock up. There is a positive correlation between mean item value – the average price a customer pays for an item in any given order – and home delivery adoption, as well as the average days between orders. The team found that high mean item value as well as higher 'avg_ord_gap' means higher home delivery adoption.




```
                            OLS Regression Results
==============================================================================
Dep. Variable:          delivery_ratio   R-squared:                       0.567
Model:                             OLS   Adj. R-squared:                  0.559
Method:                  Least Squares   F-statistic:                     70.64
Date:                 Thu, 03 May 2018   Prob (F-statistic):           2.21e-11
Time:                         23:06:36   Log-Likelihood:                 114.91
No. Observations:                   56   AIC:                            -225.8
Df Residuals:                       54   BIC:                            -221.8
Df Model:                            1
Covariance Type:             nonrobust
===========================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------
Intercept                   0.0317      0.009      3.430      0.001       0.013       0.050
mean_item_value_rounded     0.0220      0.003      8.405      0.000       0.017       0.027
==============================================================================
Omnibus:                         4.075   Durbin-Watson:                   1.079
Prob(Omnibus):                   0.130   Jarque-Bera (JB):                2.795
Skew:                            0.378   Prob(JB):                        0.247
Kurtosis:                        2.208   Cond. No.                         8.24
==============================================================================
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:          delivery_ratio   R-squared:                       0.844
Model:                             OLS   Adj. R-squared:                  0.838
Method:                  Least Squares   F-statistic:                     124.8
Date:                 Thu, 03 May 2018   Prob (F-statistic):           9.10e-11
Time:                         23:06:49   Log-Likelihood:                 80.530
No. Observations:                   25   AIC:                            -157.1
Df Residuals:                       23   BIC:                            -154.6
Df Model:                            1
Covariance Type:             nonrobust
=======================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
Intercept               0.0360      0.004      9.212      0.000       0.028       0.044
avg_ord_gap_rounded     0.0031      0.000     11.171      0.000       0.003       0.004
==============================================================================
Omnibus:                         4.081   Durbin-Watson:                   1.223
Prob(Omnibus):                   0.130   Jarque-Bera (JB):                2.745
Skew:                           -0.802   Prob(JB):                        0.253
Kurtosis:                        3.249   Cond. No.                         27.3
==============================================================================
```

Figure 4-13: Delivery Ratio as a Function of Order Frequency & Item Value

Figure 14 plot the expected 'delivery_ratio' for customers grouped by mean item value and days between orders, respectively. The above correlations explain 55.9% & 88.3% of our data, as per the adjusted R-squared of 0.559 and 0.883, respectively. The MIT team interpret these results as suggesting that there is a strong correlation between home delivery adoption rate and both average item value as well as frequency of orders.

4.1.6 Heat Map

With the model trained and tested on the customers located in regions with both fulfillment channels (home delivery and pick-up from store), the MIT team deployed the model to predict the home delivery adopters using the data of the customers located in regions with pick-up from store channel only.

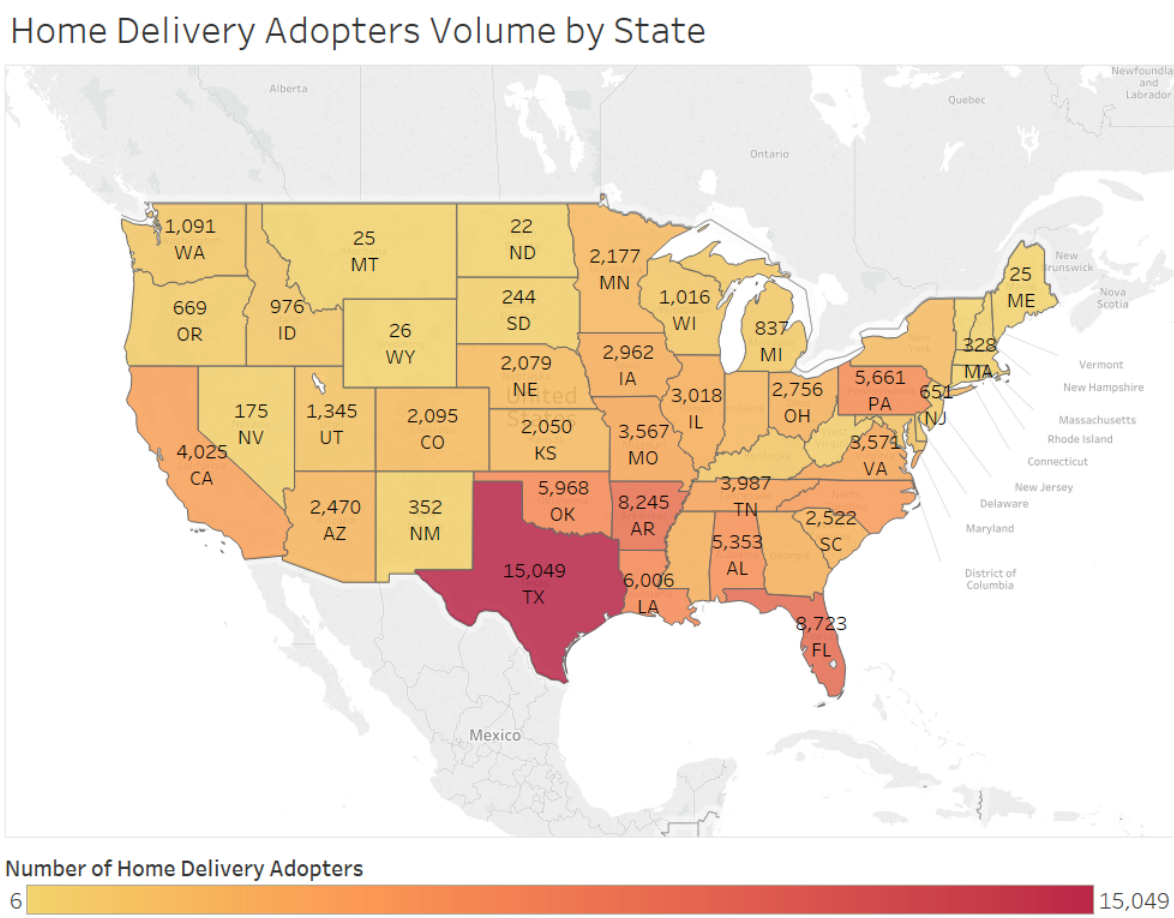


Figure 4-14: Heat Map of Volume of Home Delivery Adopters by State

The results were aggregated by zip code and tabulated in Appendix D. The zip codes were ranked by predicted number of home delivery adopters (Table D-1: Top 20 Post Codes

Ranked by Number of Home Delivery Adopters) and by predicted density of home delivery adopters (Table D-2: Top 20 Post Codes Ranked by Density of Home Delivery Adopters). The density of home delivery adopters is the ratio of predicted number of home delivery adopters over total number of Walmart.com's grocery customers in the zip code. In addition, the results were aggregated by state to build two heat maps: Figure 15 and Figure 16

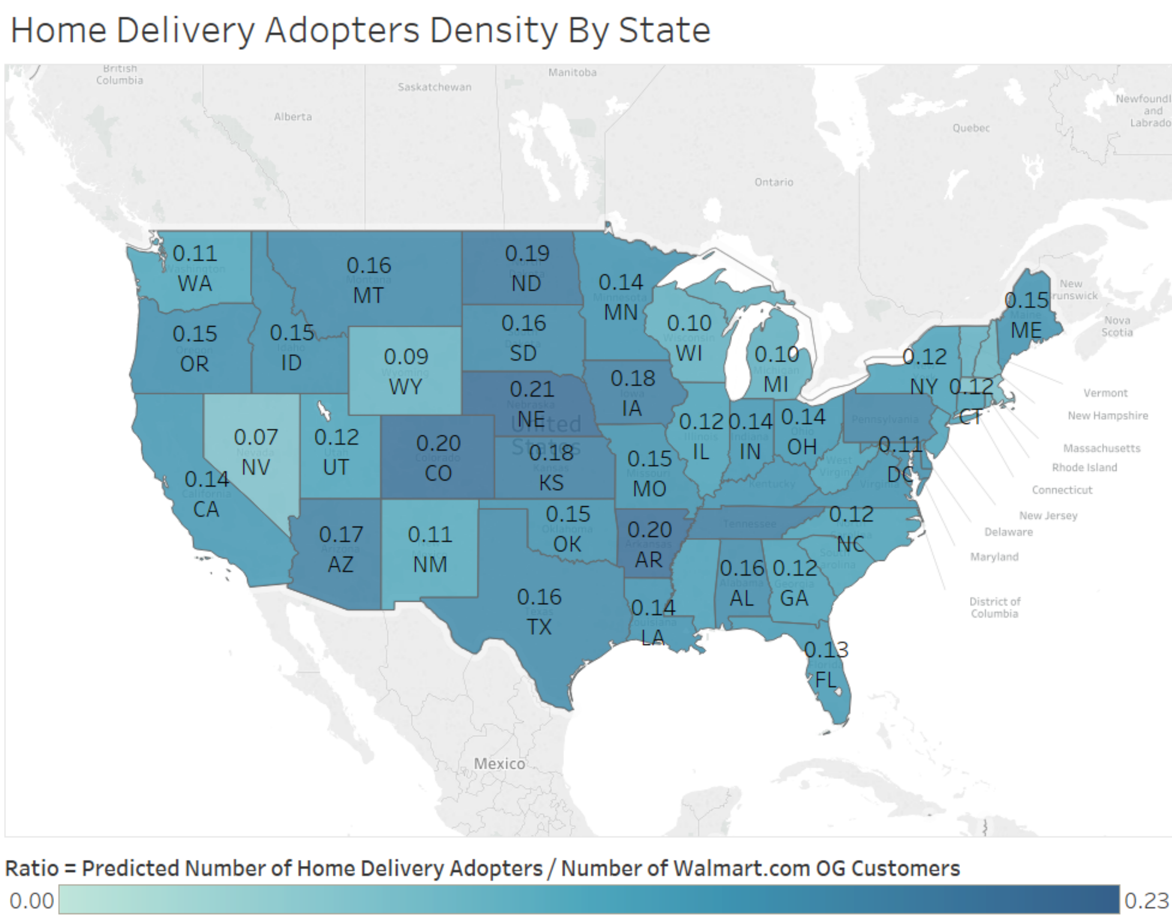


Figure 4-15: Heat Map of Density of Home Delivery Adopters by State

4.2 Channel Choice Pipeline

4.2.1 Random Effects Logit Model

The survey was released on April 16 and 801 responses were collected by April 23. Of those, 358 (44.69%) respondents identified as never having shopped for groceries online before, while 443 (55.31%) did. Demographic composition, as well as distributions of responses to all questions (including access to competitors, frequency of online grocery shopping, etc.), are included as Appendix E.

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. For the model, the 429 customers made 3,432 observations, where each customer's eight choices formed a single observation. A variety of RELM models were built to test hypotheses with the following results:

1. Customer sensitivity to price combined with delivery window was quantified such that every increase in the level of delivery fee, as described in Table 4-3, causes a person to be 20.7% less likely to choose home delivery. With a P<0.000 and a regression coefficient = -.2324817, this likelihood estimate is statistically significant and a primary driver of channel choice. While this relationship holds for across both segments (experienced

online grocery shoppers and non-experienced), it is even more prevalent among experienced online grocery shoppers.

Table 4-3: Levels of Delivery Window and Cost for Home Delivery Channel

Window	**Cost**
Order placed anytime today and delivered the next day	$6.99
Order placed by 1pm and delivered as soon as 4 hours	$9.99
Order placed by 1pm and delivered as soon as 1 hours	$14.99

2. Customer sensitivity to pick-up window was quantified such that this channel feature exhibited minimum moderating effect on channel choice. Changing the pick-up from store window from same day delivery to next day does not have change the customer's choice to the delivery channel. For the respondents who previously ordered online groceries, this channel feature's moderating coefficient =-.7857296 ($P<0.009$) negate the main effect coefficient=.7077762 ($P<0.002$)

3. 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. With a $P<0.000$ and a regression coefficient = .774097, this likelihood estimate is statistically significant and a primary driver of channel choice. Further, the analysis revealed that a customer is 1.73 times more likely to choose home delivery when a store is 10-15 miles away as compared to a customer with a store that is less than 10 miles away. With a $P<0.016$ and a regression

coefficient= .5464039, this likelihood estimate is statistically significant. That is to say that a customer becomes significantly more likely to choose home delivery as distance increases, according to the above likelihood estimates.

4. Customer sensitivity to delivery agent was quantified such that delivery agent exhibited no moderating effect on channel choice. Switching delivery agent from a Walmart associate to a 3rd-party like Uber or Lyft, with a $P<0.318$, does not have a significant moderating effect.

5. Customer sensitivity to having a car was quantified such that having a car reduces the likelihood of home delivery by 74.88%. With a $P<0.009$ and a regression coefficient= -1.381392, this likelihood estimate is statistically significant and a primary driver of channel choice. This can be interpreted as meaning that a customer having a car is highly correlated with choosing the pick-up channel.

6. 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. With a $P>.057$ and regression coefficient = -1.0118, this likelihood estimate is approaching statistical significance and a potentially significant driver of channel choice. This likelihood estimate, while falling short of a 0.05 statistical significance threshold, is further supported by a gradual decrease in p-values associated with other age groups, per Table 4-4. This gradual

decrease suggests that there is a trend that is internally consistent where increased age makes a customer become less likely to choose home delivery.

Table 4-4: Regression Coefficients and P-values per Age Range

Age Range	**Coefficient**	**P > |z|**
18-24	Base Case	Base case
25-34	0.5152606	0.377
35-44	-0.3162549	0.609
45-54	0.2027541	0.751
55-64	-0.4942015	0.463
65+	-1.141155	0.133

5. CONCLUSION

The following section provides a summary of the findings in this study and proposes management recommendations based on the most significant results. It then explores limitations and offers the MIT team's thoughts on future research and how the current study provides a foundation for extending the research into omnichannel strategy for online grocery.

5.1 Insights and Management Recommendations

- 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 provides a tool with several specific operational applications. The first is the 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. The second is the predictive model itself, which, with more data, should be able to capture more nuanced trends implicit in geographic features, as Walmart continues to expand its home delivery capabilities.

- 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.

Additionally, demographics insights from the Channel Choice Pipeline guide the grocery retailer to target the customer segment most likely to adopt grocery home delivery. The promotions on home delivery services should be directed towards Generation Z (18 – 24 years old) who are 2.75 times more likely to adopt this channel than Seniors (65+) and households without a car (approximately 4 times more likely to choose grocery home delivery than households without a car). Using the tool created via Customer Profile Pipeline, Walmart can identify the key grocery home delivery markets and design promotions targeting these groups of customers.

5.2 Limitations

The model and resulting findings were constrained in a number of ways. First, the MIT team couldn't disentangle the variables of home delivery window and price because the two were perfectly correlated. The survey was designed to reflect realistic combinations of delivery window and delivery cost. Second, because customers were only asked about three distance categories for how far a store is from the customer, no conclusions could be made about the effect of incremental distance changes (e.g. "for every mile further a customer is from a Walmart store, they are [XXX%] less likely to choose home delivery").

5.3 Future Research

The research presented in this report provides ample opportunity to pursue additional research in modeling omnichannel distribution for online grocery.

The Customer Profile pipeline was based on limited data about Walmart customers, in the relatively unchartered territory of online grocery. The data set was aggregated at the customer level, which meant that the MIT team could not explore time-series analyses of the evolution of the customer experience over time, in markets that Walmart is treating as pilot programs. Future research may be able to harness more robust data sets that will naturally reflect the growth and maturity of the online grocery market itself.

The Channel Choice Pipeline, also, was limited by the nature and scope of the market for online grocery. The MIT team-based results on hypothetical scenarios about a shopping experience that is fundamentally counterintuitive to traditional grocery shopping. As rapidly changing competitive landscapes and customer expectations shape what channels retailers offer, choice modeling surveys like the Discrete Choice Experiment (DCE) deployed for this study can explore complex trade-offs between different channel attributes, across a wide variety of channels beyond home delivery and pick-up in store.

The home delivery window and price insight should be further developed. A study, like the one conducted by Rabinovich, Sousa, Park, & Golara (2018), should explore the change in revenue and cost with the variation of delivery fees. When the delivery fee varies, there is a trade-off between the volume of orders and the revenue from the fee itself. What would be the revenue loss when the fee is increased? What would be the cost savings when fewer orders are fulfilled? And what will be the revenue from the fee itself? The objective is to maximize profits from this channel. Similarly, the MIT team believes that the grocery retailer should investigate assigning an order fee for the pick-up from store channel.

This exploratory research is the first step towards defining customer's willingness to adopt a certain delivery channel and consequently allow the grocery retailer to shape demand across a selection of channels. To that end, a deeper analysis of thresholds that define a customer's willingness to switch from one channel to another would benefit any operational strategy for omnichannel distribution.

6. BIBLIOGRAPHY

Bellini, S. (2015). A Consumer Perspective on Grocery Retailers' Differentiation. *International Business Research*, *8*(8). https://doi.org/10.5539/ibr.v8n8p129

Chintagunta, P. K., Chu, J., & Cebollada, J. (2009). *Quantifying Transaction Costs in Online/Offline Grocery Channel Choice* (SSRN Scholarly Paper No. ID 1339646). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=1339646

Droogenbroeck, E. V., & Hove, L. V. (2017). Adoption of Online Grocery Shopping: Personal or Household Characteristics? *Journal of Internet Commerce*, *16*(3), 255–286. https://doi.org/10.1080/15332861.2017.1317149

Gao, F., & Su, X. (2016). Omnichannel Retail Operations with Buy-Online-and-Pick-up-in-Store. *Management Science*, *63*(8), 2478–2492. https://doi.org/10.1287/mnsc.2016.2473

Kacen, J. J., Hess, J. D., & Kevin Chiang, W.-Y. (2013). Bricks or Clicks? Consumer Attitudes toward Traditional Stores and Online Stores. *Global Economics and Management Review*, *18*(1), 12–21. https://doi.org/10.1016/S2340-1540(13)70003-3

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 881–892. https://doi.org/10.1109/TPAMI.2002.1017616

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., … Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 3146–3154). Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf

McCullough, P. R. (2002). A User's Guide to Conjoint Analysis. *MARKETING RESEARCH MAGAZINE*, 9.

McFadden, D. L. (1973). Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*. New York: Wiley.

OmniChannel: What Were Once Luxuries Are Now Necessities. (2017). Retrieved May 14, 2018, from https://www.rsrresearch.com/research/omnichannel-what-were-once-luxuries-are-now-necessities

Vilikus, O. (2014). Optimization of Sample Size and Number of Tasks per Respondent in Conjoint Studies Using Simulated Datasets.

Orme, B. K. (2010). *Getting started with conjoint analysis: strategies for product design and pricing research* (2nd ed). Madison, WI: Research Publishers.

Rabinovich, E., Sousa, R., Park, S., & Golara, S. (2018). *Omnichannel Retailing as a Balancing Act between In-Store and Home Fulfillment* (SSRN Scholarly Paper No. ID 3113878). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=3113878

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python, 5.

Verma, R., Plaschka, G. R., Hanlon, B., Livingston, A., & Kalcher, K. (2008). Predicting customer choice in services using discrete choice analysis. *IBM Systems Journal*, *47*(1), 179–191. https://doi.org/10.1147/sj.471.0179

Yee, P. M., Dahiya, T., & Kraemer, D. (2015). DHL Customer Solutions & Innovation Represented by Matthias Heutger Senior Vice President DHL Customer Solutions & Innovation 53844 Troisdorf, Germany, 43.

Appendix A: Historical Sales Data Provided by Walmart

Table 0-1: Features Related to Customer Purchasing Behavior

Feature	Description
customer_id	Unique ID of the customer
first_trans_week	Year and week of the first online order placed by the customer
last_trans_week	Year and week of the last online order placed by the customer
ord_cnt	Count of online grocery orders since *first_trans_week*
del_ord_cnt	Count of online grocery orders fulfilled via home delivery
pickup_ord_cnt	Count of online grocery orders fulfilled via pick-up
mean_sales	Average spend in USD per online order
avg_order_gap	Average gap in days between online orders
avg_ord_size	Average number of items in one online order
min_sales	Minimum order value of customer
max_sales	Maximum online order value of customer
Post_code	Zip code of billing address of customer
City	City of billing address
State	State of billing address

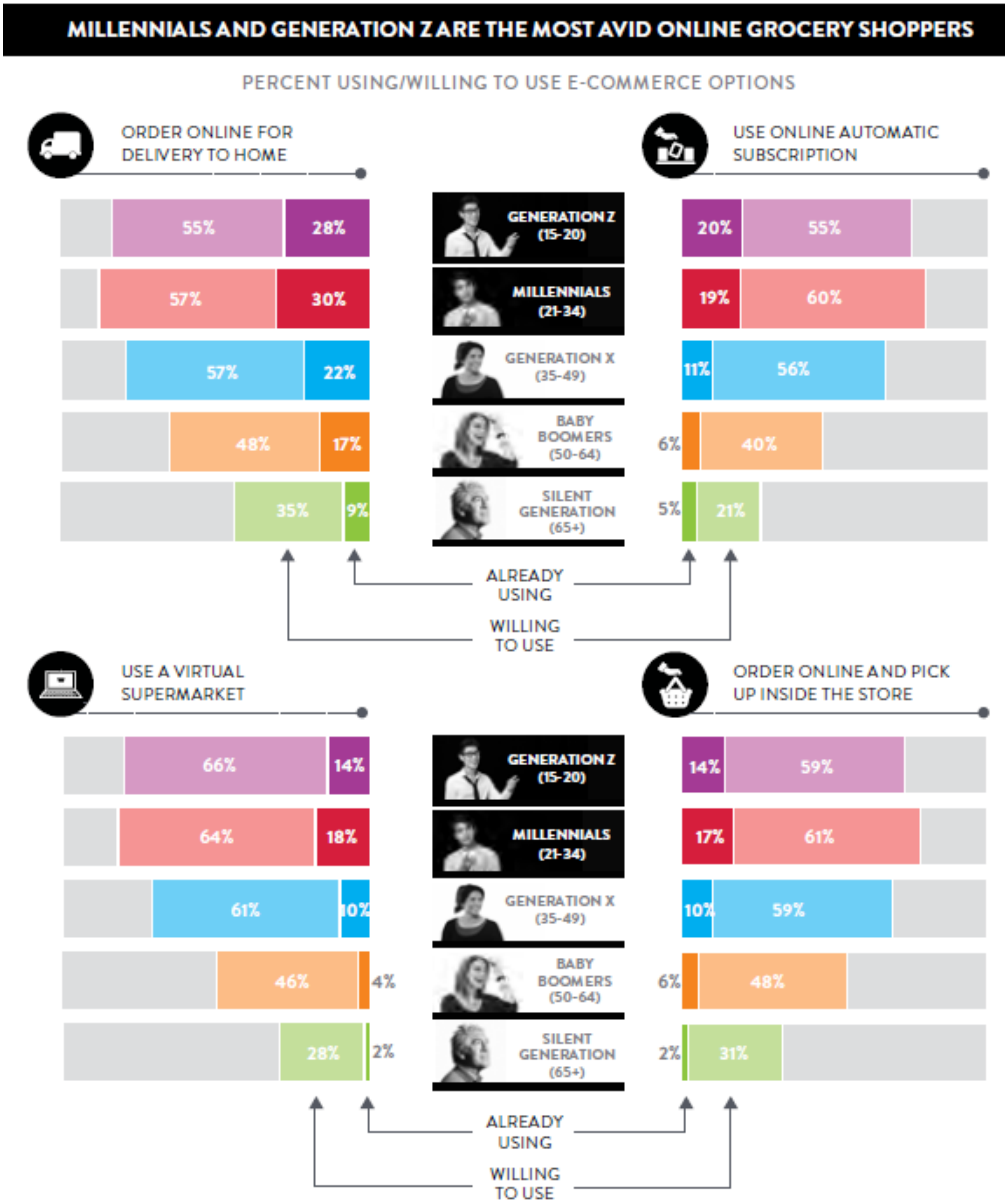
Table A-2: Features Related to Market

Feature	Description
distance miles	Distance (in miles) from zip code centroid to closest Walmart store
total comp	Total number of direct competitors in online grocery in zip code
in delivery zip	Binary variable specifying if customer has access to home delivery

Table 0A-3: Features from 2010 Census Data

Feature	Description
population	Number of inhabitants in zip code
land area	Area covered by zip code in square miles

Appendix B: Online Grocery Nielsen Survey (2014)



Source: Nielsen Global E-commerce and the New Retail Survey, Q3 2014

Appendix C: Confusion Matrices

Gradient Boosting Machine

		Predicted		
		0	**1**	Total
Actual	**0**	8,428	568	8,996
	1	689	978	1,667
	Total	9,117	1,546	

	Precision	**Recall**	**F1 score**
0	0.92	0.94	0.93
1	0.63	0.59	0.61
		Average F1 score	**0.88**

K Nearest Neighbors

		Predicted		
		0	**1**	Total
Actual	**0**	8,580	416	8,996
	1	915	752	1,667
	Total	9,495	1,168	

	Precision	**Recall**	**F1 score**
0	0.90	0.95	0.93
1	0.64	0.45	0.53
		Average F1 score	**0.87**

Naïve Bayes Classifier

		Predicted		
		0	**1**	Total
Actual	**0**	8,638	358	8,996
	1	927	740	1,667
	Total	9,565	1,098	

	Precision	**Recall**	**F1 score**
0	0.90	0.96	0.93
1	0.67	0.44	0.54
		Average F1 score	**0.87**

Appendix D: Predictions of Home Delivery Adopters

Table 0-1: Top 20 Post Codes Ranked By Number of Home Delivery Adopters

	Post Code	City	State	Predicted Number of Home Delivery Adopters	Predicted Density of Home Delivery Adopters
1	72712	Bentonville	AR	893	0.26
2	72758	Rogers	AR	824	0.26
3	73034	Edmond	OK	513	0.24
4	39564	Ocean Springs	MS	407	0.2
5	72762	Springdale	AR	387	0.19
6	72756	Rogers	AR	380	0.28
7	72703	Fayetteville	AR	330	0.2
8	72764	Springdale	AR	287	0.25
9	84074	Tooele	UT	279	0.13
10	78130	New Braunfels	TX	275	0.22
11	39565	Vancleave	MS	271	0.28
12	72719	Centerton	AR	263	0.32
13	72956	Van Buren	AR	259	0.16
14	72714	Bella Vista	AR	257	0.36
15	73013	Edmond	OK	247	0.1
16	80831	Peyton	CO	247	0.22
17	72701	Fayetteville	AR	237	0.14
18	35613	Athens	AL	233	0.28
19	72745	Lowell	AR	226	0.25
20	37334	Fayetteville	TN	222	0.42

Table 0-2: Top 20 Post Codes Ranked By Density of Home Delivery Adopters

	Post Code	City	State	Predicted Number of Home Delivery Adopters	Predicted Density of Home Delivery Adopters
1	94040	Mountain View	CA	36	0.73
2	35671	Tanner	AL	15	0.71
3	66523	Osage City	KS	13	0.68
4	80832	Ramah	CO	12	0.67
5	99029	Reardan	WA	17	0.65
6	38478	Pulaski	TN	11	0.65
7	76673	Mount Calm	TX	13	0.62
8	79358	Ropesville	TX	13	0.62
9	16050	Petrolia	PA	19	0.61
10	78056	Mico	TX	15	0.6
11	76431	Chico	TX	12	0.57
12	80835	Simla	CO	12	0.57
13	68358	Firth	NE	19	0.56
14	80808	Calhan	CO	86	0.55
15	37058	Dover	TN	62	0.55
16	80135	Sedalia	CO	20	0.54
17	37308	Birchwood	TN	15	0.54
18	78662	Red Rock	TX	15	0.54
19	81624	Collbran	CO	14	0.54
20	81523	Glade Park	CO	13	0.54

Appendix E: Survey Response Distributions

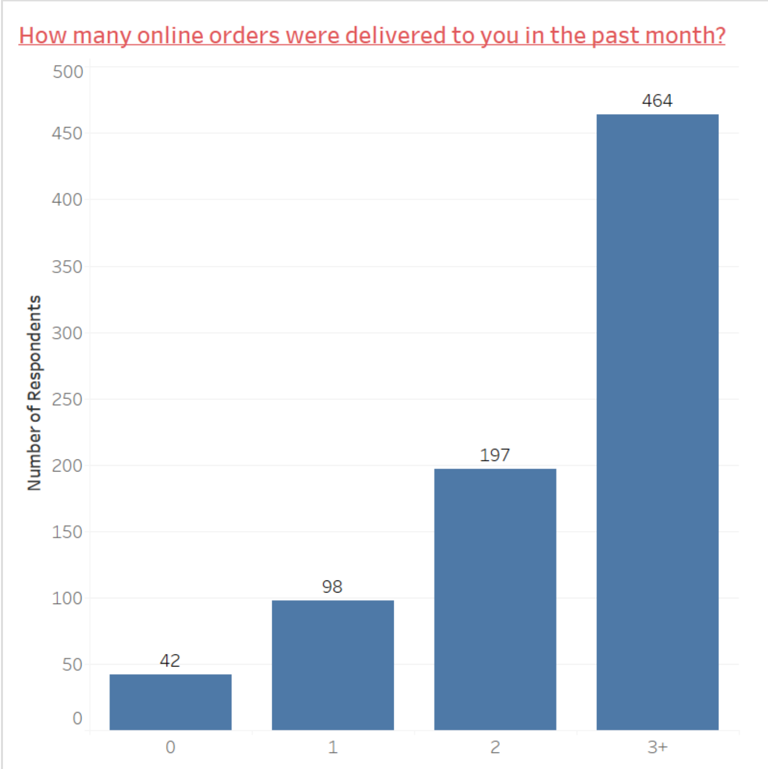


Figure E-1: Survey Answers – Online Order Frequency

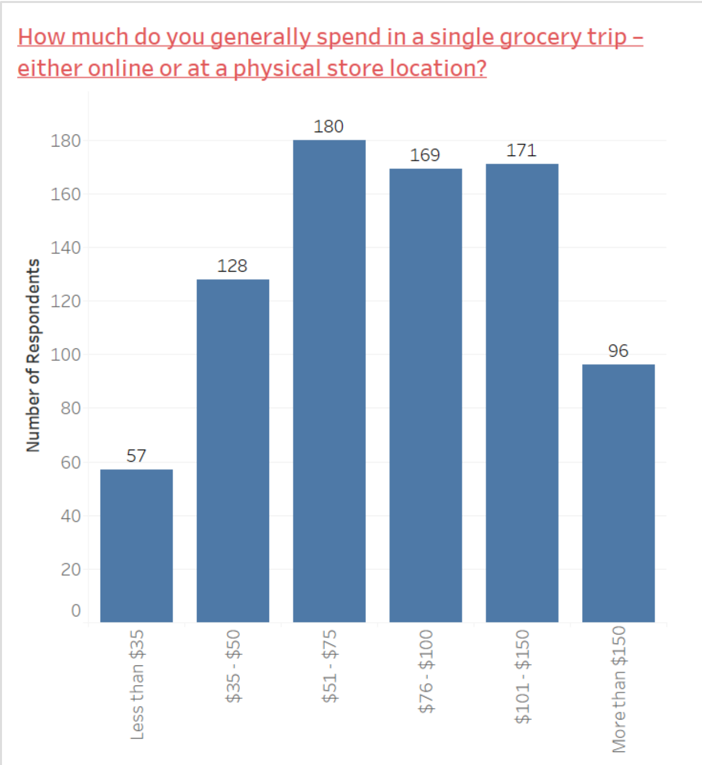


Figure E-2: Survey Answers – Average Spend for Groceries

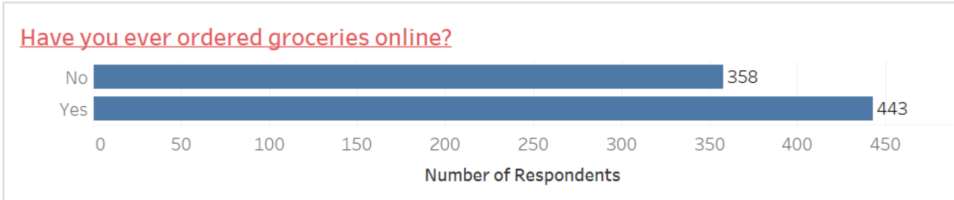


Figure E-3: Survey Answers – Users of Online Grocery

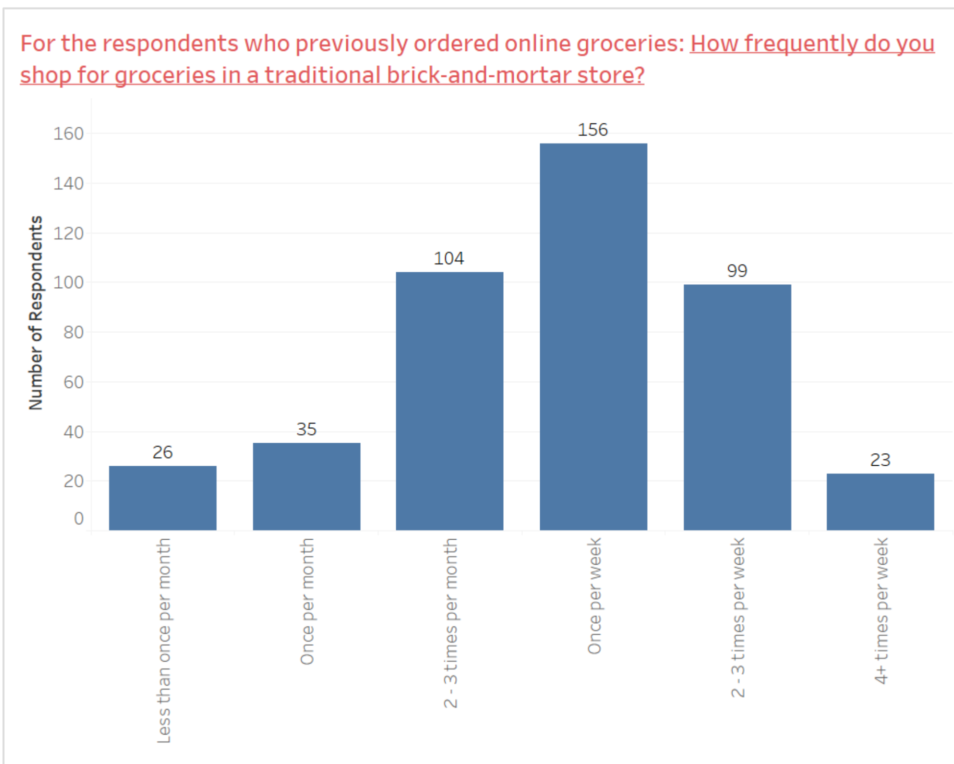


Figure E-4: Survey Answers – Frequency of in-Store Grocery Shopping

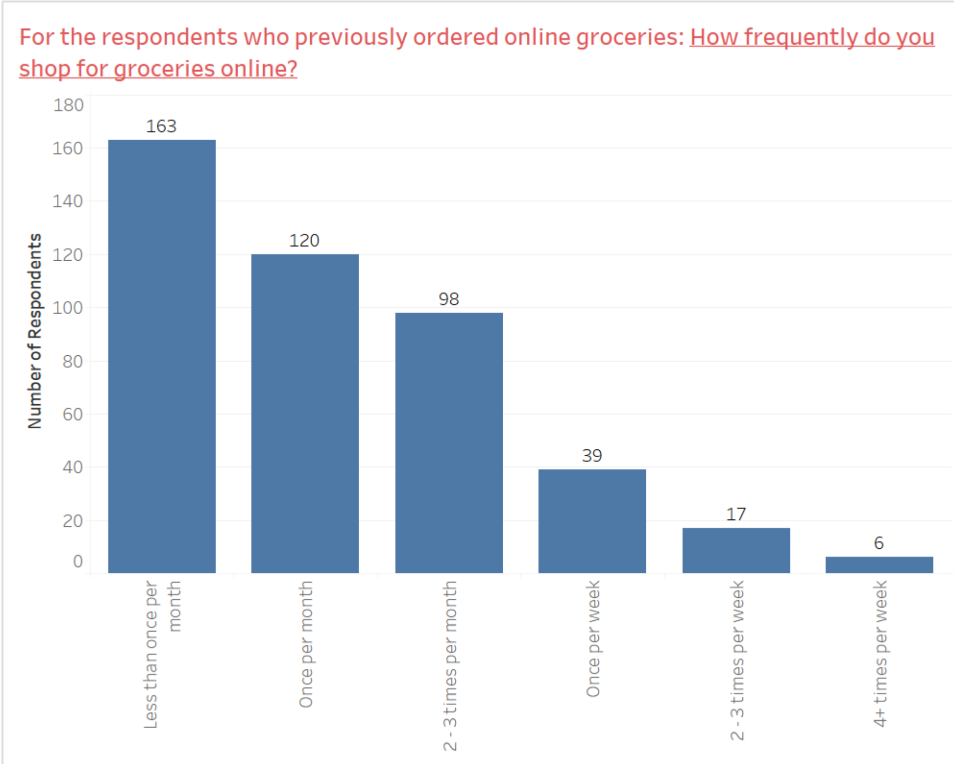


Figure E-5: Survey Answers – Online Grocery Frequency

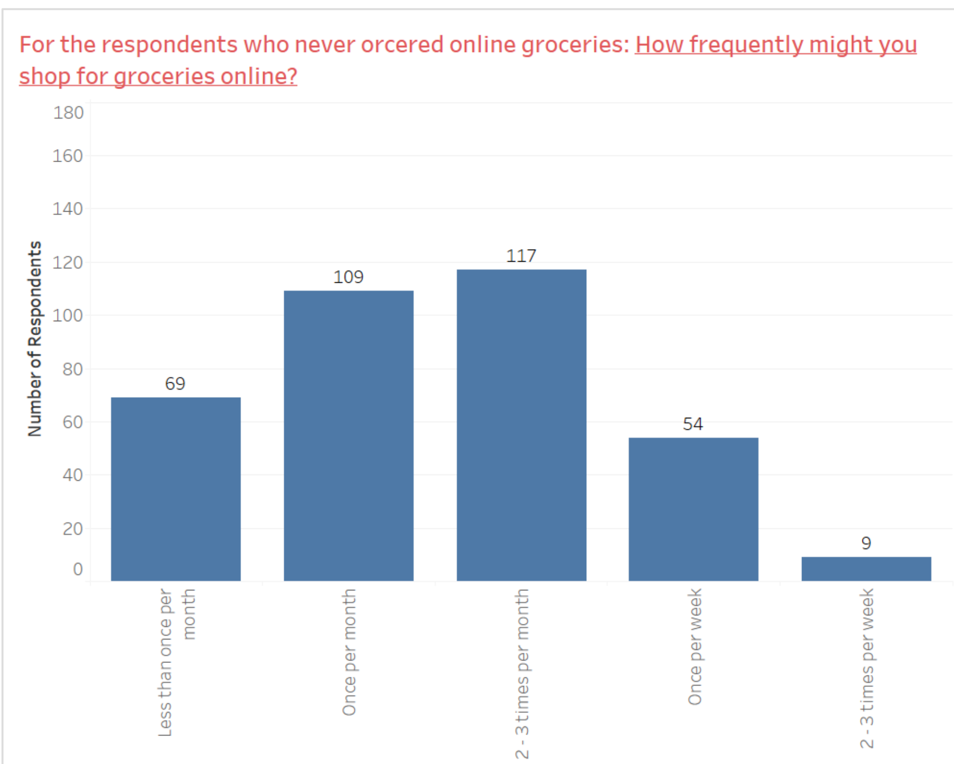


Figure E-6: Survey Answers – Expected Online Grocery Frequency

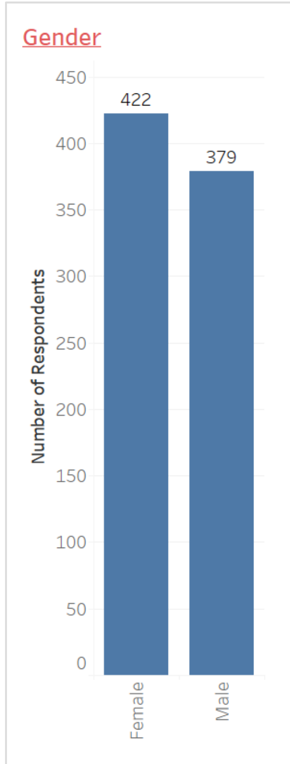


Figure 0E-7: Survey Respondents' Demographics – Gender

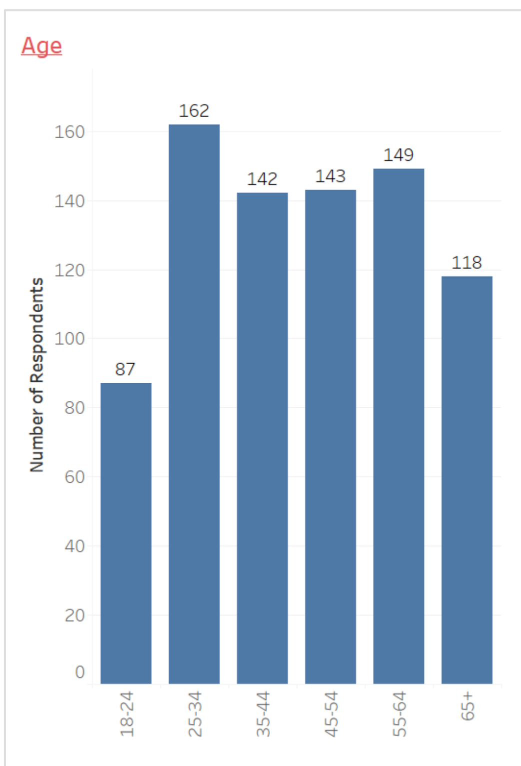


Figure E-8: Survey Respondents' Demographics – Age

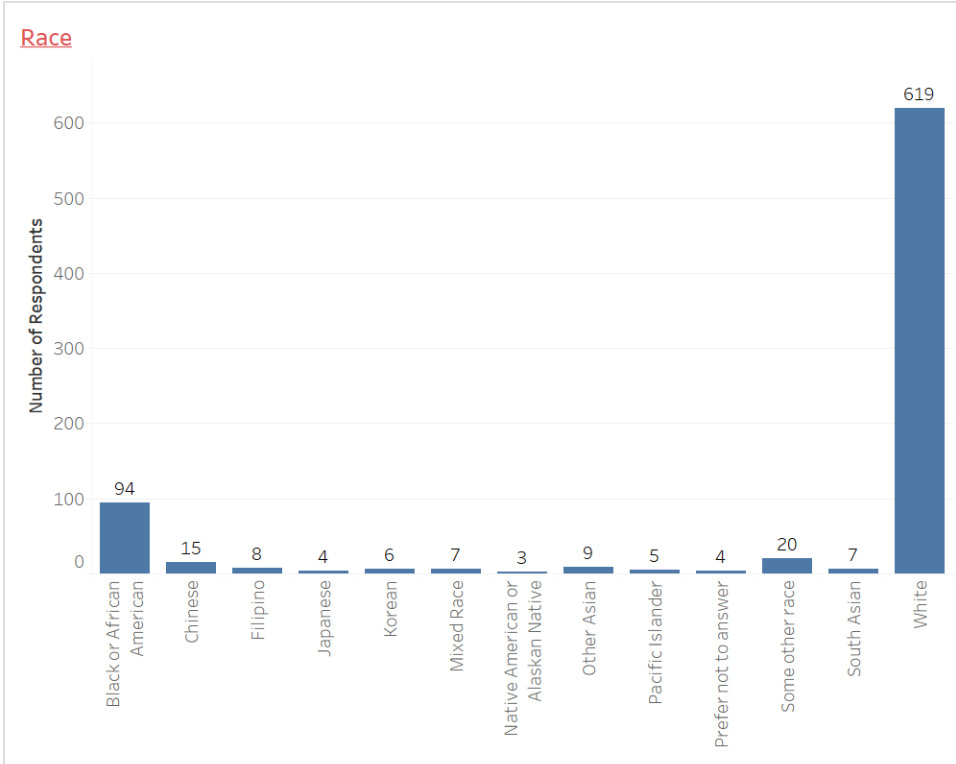


Figure E-9: Survey Respondents' Demographics – Race

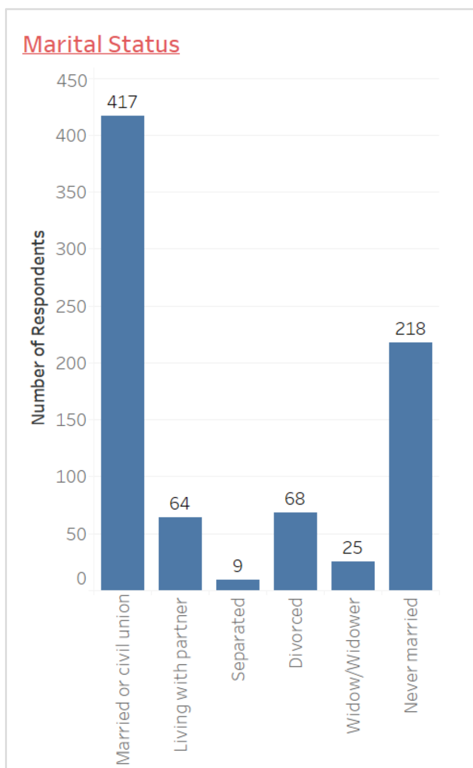


Figure E-10: Survey Respondents' Demographics – Marital Status

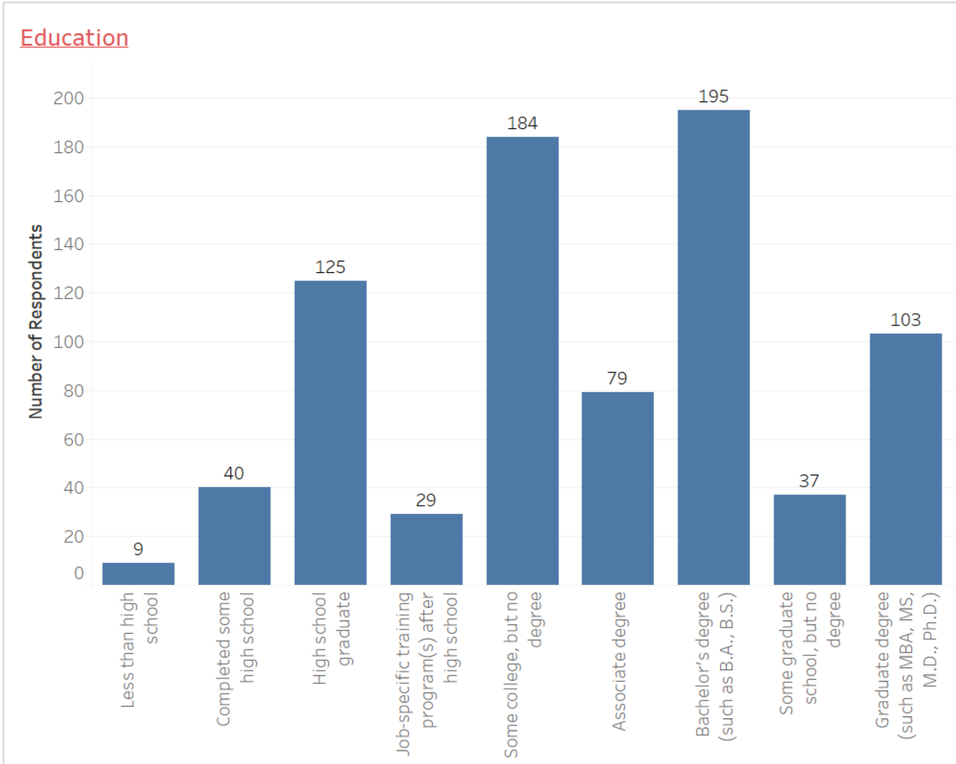


Figure E-11: Survey Respondents' Demographics – Education

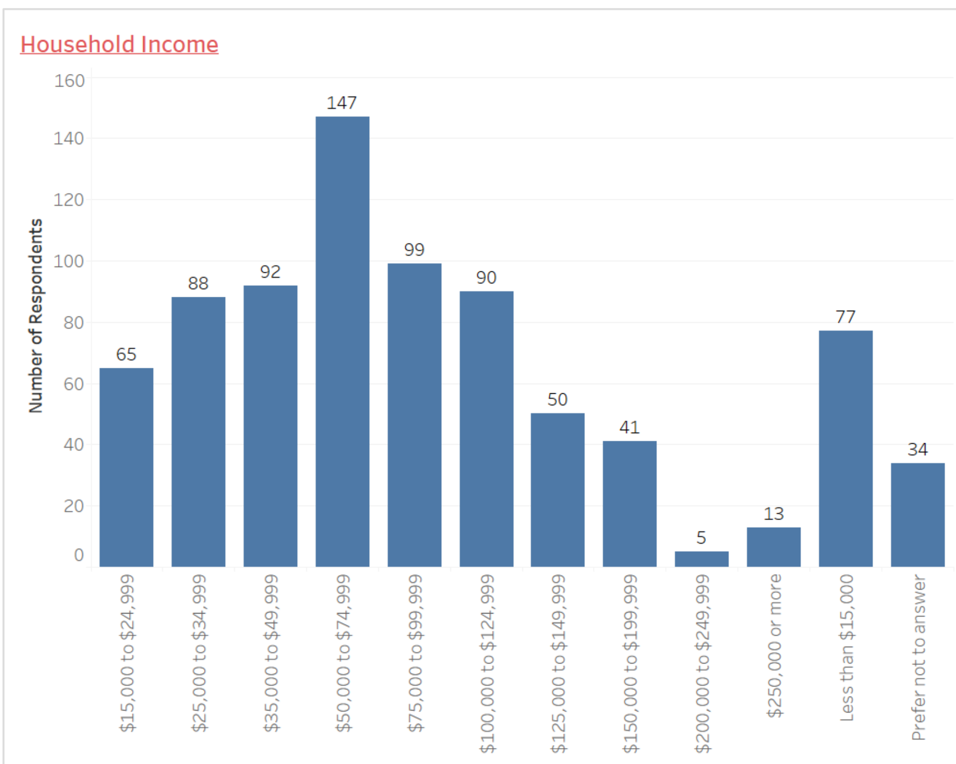


Figure 0E-12: Survey Respondents' Demographics – Household Income

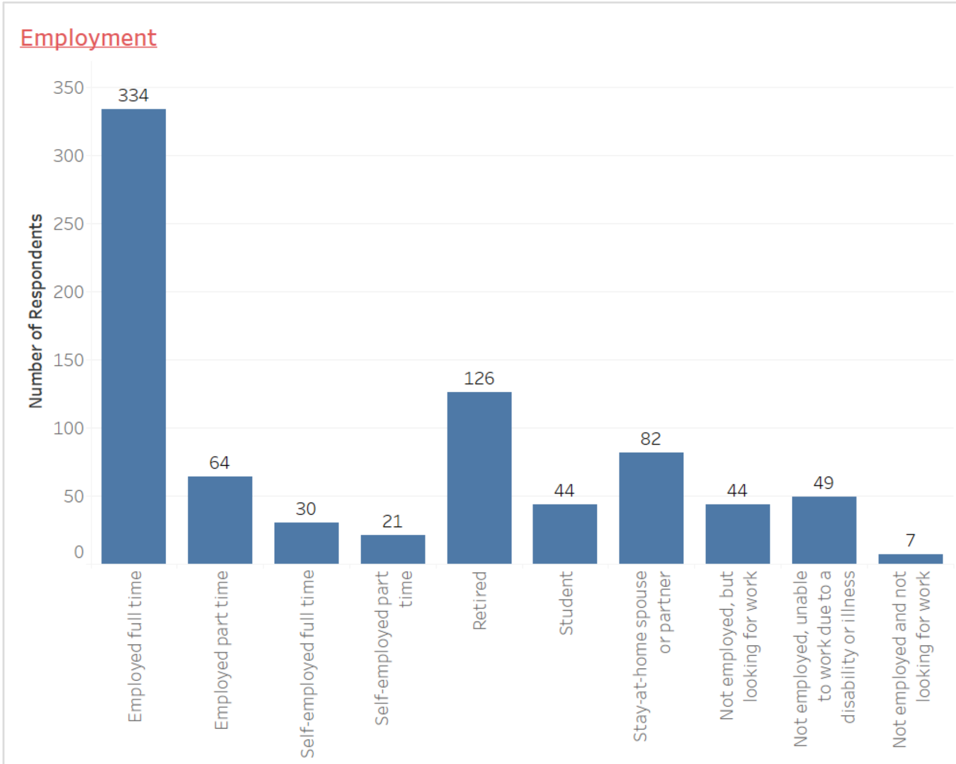


Figure E-13: Survey Respondents' Demographics – Employment

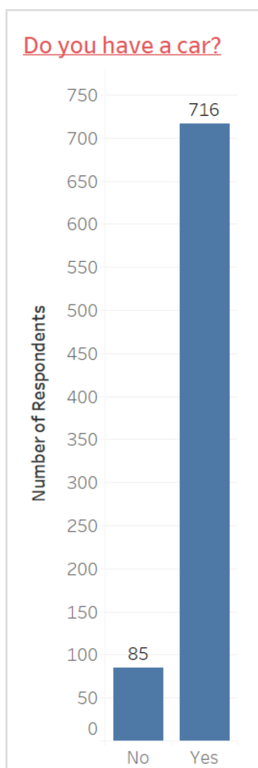


Figure E-14: Survey Answers – Car Availability

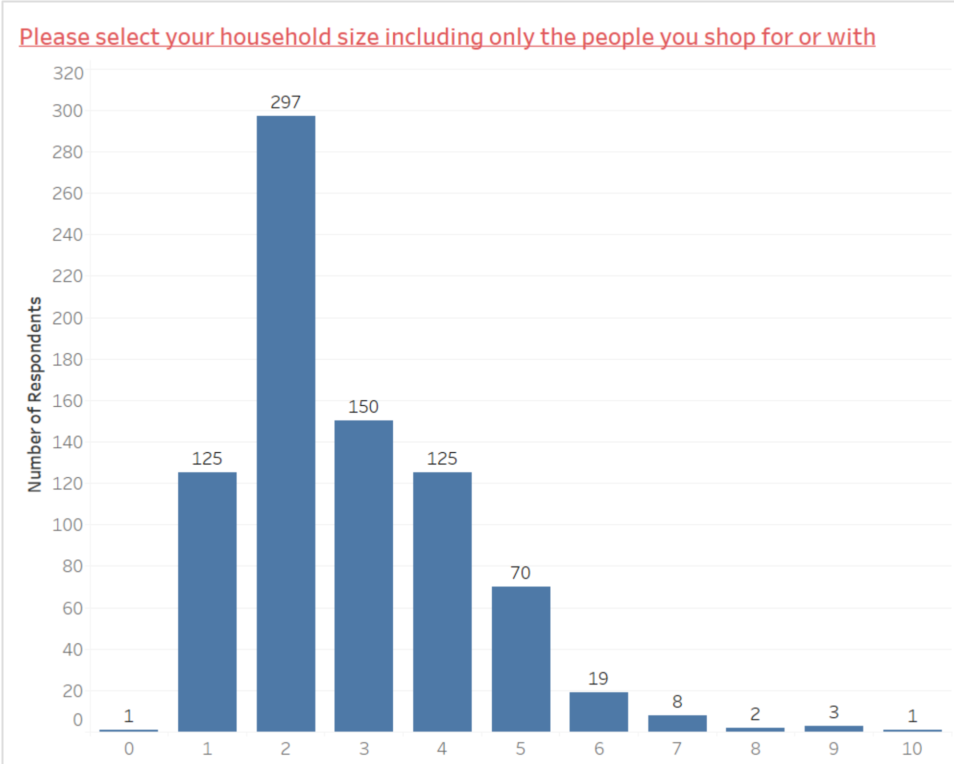


Figure E-15: Survey Respondents' Demographics – Household Size

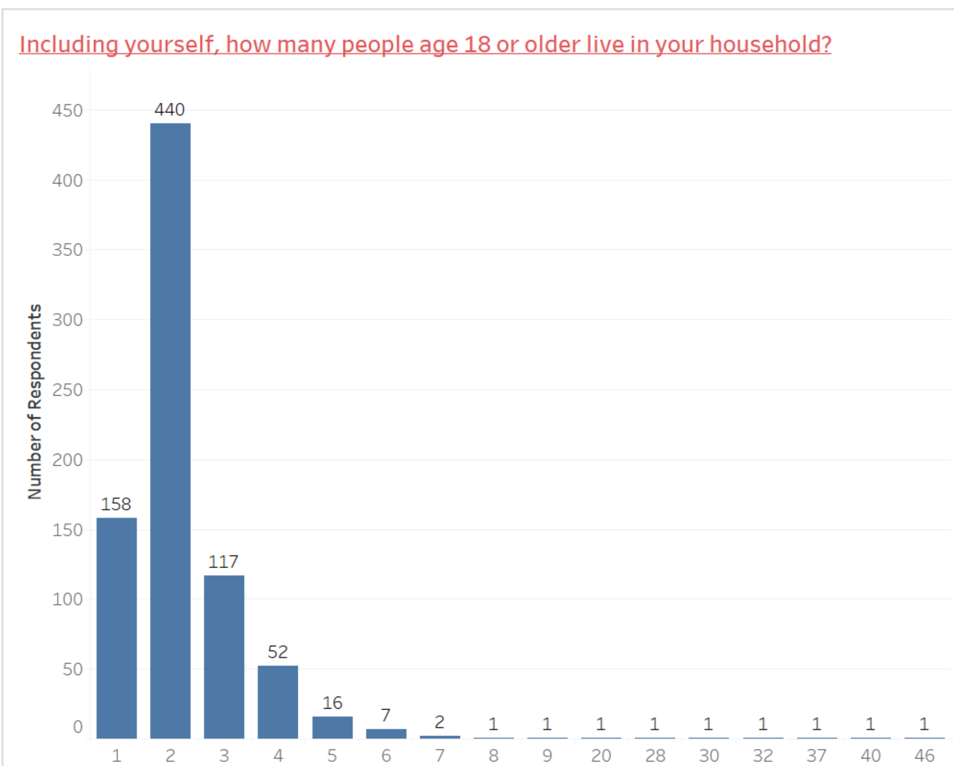


Figure E-16: Survey Respondents' Demographics – Household Size (Adults)

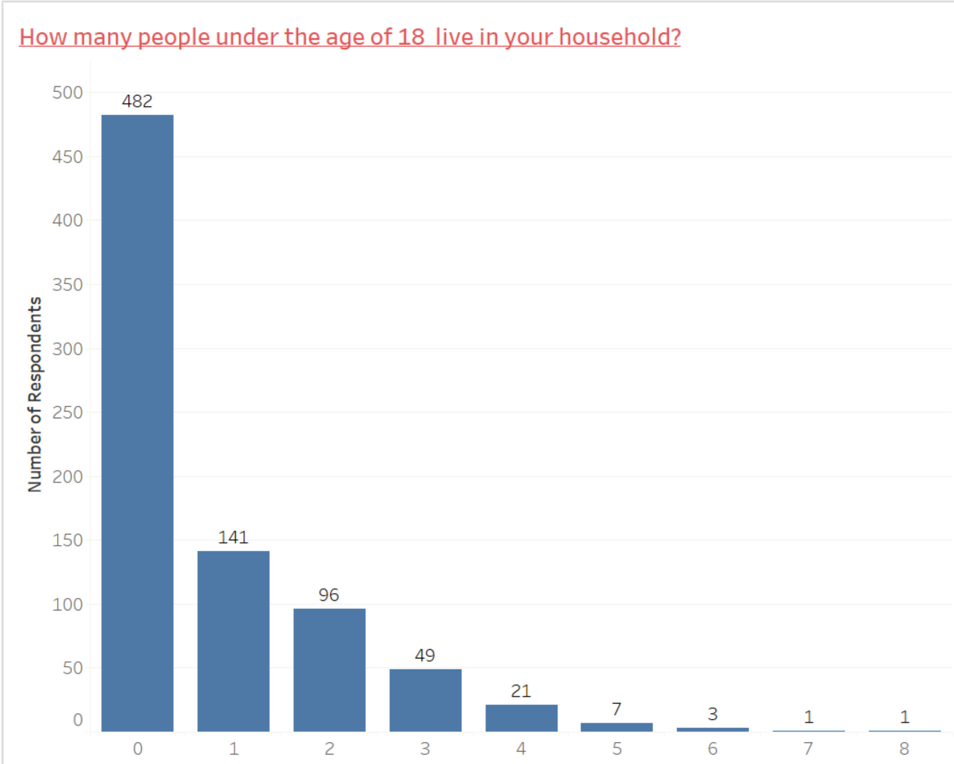


Figure E-17: Survey Respondents' Demographics – Household Size (Minors)

Table 0-1: Users of Online grocery Retailers Among Survey Respondents

Online Grocery Retailers	Number of Users
Amazon Prime Now	176
Walmart Online Grocery	146
Amazon Fresh	136
Jet.com	45
Peapod	38
Fresh Direct	31
Instacart	31
Krogers Clicklist	31
Safeway	25
Harris Teeter	11
Fred Meyer	10
Fry's Foods	9
ShopRite	7
Mariano's	5
Smith's	5

Appendix F: Survey's Logic, Questions and Choice Sets

Job Name:	[Walmart Grocery]	
Landing Page Title (scrIntroTitle)	Thank you for taking our survey.	
Job No. (dmJobNum)		**P140482**
LOI for ISQ section (isqLOI)		[10] (minutes)
Digital Fingerprinting and Fraud Score [Imperium RelevantID] (dfOptions) By default, surveys will terminate any respondents who fail both of these tests and is recommended for panel sample. For client sample or vendor sample, the termination of DF or Fraud Score can be turned off if desired. *[SELECT CODES 1-3]*		1 Digital Fingerprinting/Fraud Score 2 Terminate DF Duplicate 3 Terminate DF Fraud Score
Mode of survey (srvyMode)		**1 -** Web **2 -** CATI/COW
Sample Sources List sample sources used for the study; Client, HPOL etc.		
Other notes OR use for client summary		None

SAMPLE PRELOAD AND SCREENING QUESTIONS

BASE: ALL RESPONDENTS
scrIntroTitle [Thank you for taking our survey.]

scrIntro
During the survey, please do not use your browser's FORWARD and BACK buttons. Instead, please always use the button below to move through the survey. Please be aware that once you've answered a question, you might not be able to go back and change your answer.

The progress bar below indicates approximately what portion of the survey you have completed.

Simply click on the button at the bottom of the page to begin the survey.

BASE: ALL RESPONDENTS
dmCntry [Country] In which country or region do you currently reside?

[IF US STUDY, US LISTED FIRST ELSE CODES DISPLAYED IN ALPHABETICAL ORDER]

244. United States of America
14. Australia
33. Brazil
42. Canada
48. China
76. France
85. Germany
116. India
123. Italy
126. Japan
157. Mexico
196. Russian Federation
215. Spain

243. United Kingdom
996. Other country

[PN: IF NOT dmCntry/244 US TERMINATE IMMEDIATELY.]

[dmGen, dmAge PRESENTED ON SAME SCREEN.]

BASE: ALL RESPONDENTS
dmGen [Gender] What is your gender?

1. Male
2. Female

BASE: ALL RESPONDENTS
dmAge [Age] What is your age?

Please enter a numeric response only.

[RANGE 0 -120]

|__|__|__|

[PN: IF NOT 18+ TERMINATE IMMEDIATELY.]

UPDATE netAge AND netGenAge TO MATCH THE WEIGHTING TARGETS.

BASE: ALL RESPONDENTS
netAge HIDDEN: Age (Net)

[COMPUTE AGE FROM dmAge]

1. 18-24
2. 25-34
3. 35-44
4. 45-54
5. 55-64
6. 65+

BASE: ALL RESPONDENTS
netGenAge HIDDEN: Gender Age (Net)

[COMPUTE AGE/GENDER FROM dmGen AND dmAge]

1. Male 18-24
2. Male 25-34
3. Male 35-44
4. Male 45-54
5. Male 55-64
6. Male 65+
7. Female 18-24
8. Female 25-34

9. Female 35-44
10. Female 45-54
11. Female 55-64
12. Female 65+

BASE: ALL RESPONDENTS
1. How many online orders were delivered to you in the past month?

This could be for any type of product and from any online channel.

1. 0
2. 1
3. 2
4. 3+

BASE: ALL RESPONDENTS
2. Have you ever ordered groceries online before today?

For groceries, please consider Fresh/Perishables (produce, meat, dairy, frozen foods, etc.) and/or Pantry items (non-perishable items, pet food, cereal, etc.).

1. Yes
2. No

BASE: RESPONDENTS WHO HAVE NOT ORDERED ONLINE (Q2/2)
3. Would you be interested in ordering groceries online?

1. Yes
2. No

QUOTAS

Qualification: US, Age 18+, Have Ordered Groceries Online (Q2/1) OR are Interested (Q3/1)
N=800

MAIN SURVEY

BASE: RESPONDENTS WHO HAVE NOT ORDERED GROCERIES ONLINE (Q2/2), BUT ARE INTERESTED (Q3/1)
4. Which of the following online grocery providers are available to you, in your area?

Please select all that apply.

[RANDOMIZE, MULTIPLE CHOICE]

1. Walmart Online Grocery
2. Jet.com
3. Instacart

4. Amazon Fresh
5. Amazon Prime Now
6. Krogers Clicklist
7. Fred Meyer
8. Fry's Foods
9. Harris Teeter
10. Smith's
11. Mariano's
12. Peapod
13. Fresh Direct
14. Safeway
15. Other: [TEXT BOX] END ANCHOR

BASE: RESPONDENTS WHO HAVE NOT ORDERED GROCERIES ONLINE (Q2/2), BUT ARE INTERESTED (Q3/1)

5. How frequently might you shop for groceries online?

1. Less than once per month
2. Once per month
3. 2 – 3 times per month
4. Once per week
5. 2 – 3 times per week

BASE: RESPONDENTS WHO ORDERED GROCERIES ONLINE (Q2/1)

6. From which of the following have you ordered groceries online?

Please select all that apply.

[RANDOMIZE, MULTIPLE CHOICE]

1. Walmart Online Grocery
2. Jet.com
3. Instacart
4. Amazon Fresh
5. Amazon Prime Now
6. Krogers Clicklist
7. Fred Meyer
8. Fry's Foods
9. Harris Teeter
10. Smith's
11. Mariano's
12. Peapod
13. Fresh Direct
14. Safeway
15. Other: [TEXT BOX] END ANCHOR

[PN: SHOW Q7 AND Q8 ON ONE SCREEN; RANDOMIZE POSITION]

BASE: RESPONDENTS WHO ORDERED GROCERIES ONLINE (Q2/1)

7. How frequently do you shop for groceries online?

1. Less than once per month
2. Once per month
3. 2 – 3 times per month
4. Once per week
5. 2 – 3 times per week
6. 4+ times per week

BASE: RESPONDENTS WHO ORDERED GROCERIES ONLINE (Q2/1)

8. How frequently do you shop for groceries in a traditional brick-and-mortar store?

1. Less than once per month
2. Once per month
3. 2 – 3 times per month
4. Once per week
5. 2 – 3 times per week
6. 4+ times per week

BASE: ALL QUALIFIED RESPONDENTS

10. How much do you generally spend in a single grocery trip – either online or at a physical store location?

1. Less than $35
2. $35 – $50
3. $51 – $75
4. $76 – $100
5. $101 – $150
6. More than $150

BASE: ALL QUALIFIED RESPONDENTS

9.We would like you to now imagine you are about to complete your purchase of groceries online through Walmart Online Grocery.

We are going to show you several different online purchase scenarios, each with two options. For each scenario, please review the options and then select the one you would choose.

Again, please consider groceries to be Frozen/Fresh/Perishable items (produce, meat, dairy, etc.) and/or Pantry/non-perishables (pet food, cereal, etc.). Also, consider it has to be a minimum $30 grocery purchase.

[PN: SHOW ALL EIGHT CONCEPTS; RANDOMIZE ORDER; SHOW COUNTER]

Choice Set 1

We would like you to now imagine you are about to complete your purchase of groceries online - through Walmart Online Grocery. We are going to show you several different online purchase scenarios, each with two options. For each scenario, please review the options and then select the one you would choose.

Groceries are Fresh/Perishables items (frozen, produce, meat, dairy etc.) and/or Pantry items (consumables, non-perishable items, pet food etc.)

	Home Delivery	Pick Up In Store
	Order placed by 1pm and delivered as soon as 4 hours	Order placed by 1pm and ready for pick-up as soon as 4 hours
	Order delivered by Walmart Associate	Curbside Pick-up
	Delivery fee is $9.99	Free
	Store is more than 15 miles from home	Store is more than 15 miles from home
	Store is not on your daily commute	Store is not on your daily commute
	You must be present during delivery window	
Which option would you choose?	☐	☐

Choice Set 2

We would like you to now imagine you are about to complete your purchase of groceries online - through Walmart Online Grocery. We are going to show you several different online purchase scenarios, each with two options. For each scenario, please review the options and then select the one you would choose.

Groceries are Fresh/Perishables items (frozen, produce, meat, dairy etc.) and/or Pantry items (consumables, non-perishable items, pet food etc.)

	Home Delivery	Pick Up In Store
	Order placed by 1pm and delivered as soon as 4 hours	Order placed by 1pm and ready for pick-up as soon as 4 hours
	Order delivered by 3rd party (e.g. Uber/Deliv)	Curbside Pick-up
	Delivery fee is $9.99	Free
	Store is more than 15 miles from home	Store is more than 15 miles from home
	Store is not on your daily commute	Store is not on your daily commute
	You must be present during delivery window	
Which option would you choose?	☐	☐

Choice Set 3

We would like you to now imagine you are about to complete your purchase of groceries online - through Walmart Online Grocery. We are going to show you several different online purchase scenarios, each with two options. For each scenario, please review the options and then select the one you would choose.

Groceries are Fresh/Perishables items (frozen, produce, meat, dairy etc.) and/or Pantry items (consumables, non-perishable items, pet food etc.)

	Home Delivery	Pick Up In Store
	Order placed by 1pm and delivered as soon as one hours	Order placed by 1pm and ready for pick-up as soon as 4 hours
	Order delivered by Walmart Associate	Curbside Pick-up
	Delivery fee is $14.99	Free
	Store is between 10 and 15 miles from home	Store is between 10 and 15 miles from home
	Store is on your daily commute	Store is on your daily commute
	You must be present during delivery window	
Which option would you choose?	☐	☐

Choice Set 4

We would like you to now imagine you are about to complete your purchase of groceries online - through Walmart Online Grocery. We are going to show you several different online purchase scenarios, each with two options. For each scenario, please review the options and then select the one you would choose.

Groceries are Fresh/Perishables items (frozen, produce, meat, dairy etc.) and/or Pantry items (consumables, non-perishable items, pet food etc.)

	Home Delivery	Pick Up In Store
	Order placed anytime today and delivered next day	Order placed by 1pm and ready for pick-up as soon as 4 hours
	Order delivered by 3rd party (e.g. Uber/Deliv)	Curbside Pick-up
	Delivery fee is $6.99	Free
	Store is less than 10 miles from home	Store is less than 10 miles from home
	Store is not on your daily commute	Store is not on your daily commute
	You must be present during delivery window	
Which option would you choose?	☐	☐

Choice Set 5

We would like you to now imagine you are about to complete your purchase of groceries online - through Walmart Online Grocery. We are going to show you several different online purchase scenarios, each with two options. For each scenario, please review the options and then select the one you would choose.

Groceries are Fresh/Perishables items (frozen, produce, meat, dairy etc.) and/or Pantry items (consumables, non-perishable items, pet food etc.)

	Home Delivery	Pick Up In Store
	Order placed anytime today and delivered next day	Order placed by 1pm and picked-up the next day
	Order delivered by 3rd party (e.g. Uber/Deliv)	Curbside Pick-up
	Delivery fee is $6.99	Free
	Store is between 10 and 15 miles from home	Store is between 10 and 15 miles from home
	Store is not on your daily commute	Store is not on your daily commute
	You must be present during delivery window	
Which option would you choose?	☐	☐

Choice Set 6

We would like you to now imagine you are about to complete your purchase of groceries online - through Walmart Online Grocery. We are going to show you several different online purchase scenarios, each with two options. For each scenario, please review the options and then select the one you would choose.

Groceries are Fresh/Perishables items (frozen, produce, meat, dairy etc.) and/or Pantry items (consumables, non-perishable items, pet food etc.)

	Home Delivery	Pick Up In Store
	Order placed by 1pm and delivered as soon as 4 hours	Order placed by 1pm and ready for pickup as soon as 4 hours
	Order delivered by Walmart Associate	Curbside pickup
	Delivery Fee is $9.99	Free
	Store is between 10 and 15 miles from home	Store is between 10 and 15 miles from home
	Store is on your daily commute	Store is on your daily commute
	You must be present during delivery window	
Which option would you choose?	☐	☐

Choice Set 7

We would like you to now imagine you are about to complete your purchase of groceries online - through Walmart Online Grocery. We are going to show you several different online purchase scenarios, each with two options. For each scenario, please review the options and then select the one you would choose.

Groceries are Fresh/Perishables items (frozen, produce, meat, dairy etc.) and/or Pantry items (consumables, non-perishable items, pet food etc.)

	Home Delivery	Pick Up In Store
	Order placed by 1pm and delivered as soon as 4 hours	Order placed by 1pm and ready for pickup as soon as 4 hours
	Order delivered by 3rd Party (e.g., Uber/Deliv)	Curbside pickup
	Delivery Fee is $9.99	Free
	Store is between 10 and 15 miles from home	Store is between 10 and 15 miles from home
	Store is not on your daily commute	Store is not on your daily commute
	You must be present during delivery window	
Which option would you choose?	☐	☐

Choice Set 8

We would like you to now imagine you are about to complete your purchase of groceries online - through Walmart Online Grocery. We are going to show you several different online purchase scenarios, each with two options. For each scenario, please review the options and then select the one you would choose.

Groceries are Fresh/Perishables items (frozen, produce, meat, dairy etc.) and/or Pantry items (consumables, non-perishable items, pet food etc.)

	Home Delivery	Pick Up In Store
	Order placed by 1pm and delivered as soon as 1 hours	Order placed by 1pm and picked up next day
	Order delivered by Walmart Associate	Curbside pickup
	Delivery Fee is $14.99	Free
	Store is between 10 and 15 miles from home	Store is between 10 and 15 miles from home
	Store is not on your daily commute	Store is not on your daily commute
	You must be present during delivery window	
Which option would you choose?	☐	☐

SECTION:	DEMOGRAPHICS

We now have a final few questions for classification purposes.

BASE: ALL US RESPONDENTS 18+
dmStateUS [State (US)] In what state or territory do you currently reside?

[DISPLAY IN DROP DOWN LIST]

1. Alabama
2. Alaska
3. Arizona
4. Arkansas
5. California
6. Colorado
7. Connecticut
8. Delaware
9. District of Columbia
10. Florida
11. Georgia
12. Hawaii
13. Idaho
14. Illinois
15. Indiana
16. Iowa
17. Kansas
18. Kentucky
19. Louisiana
20. Maine
21. Maryland
22. Massachusetts
23. Michigan
24. Minnesota
25. Mississippi
26. Missouri
27. Montana
28. Nebraska
29. Nevada
30. New Hampshire
31. New Jersey
32. New Mexico
33. New York
34. North Carolina
35. North Dakota
36. Ohio
37. Oklahoma
38. Oregon
39. Pennsylvania
40. Rhode Island
41. South Carolina
42. South Dakota
43. Tennessee
44. Texas
45. Utah
46. Vermont
47. Virginia
48. Washington
49. West Virginia
50. Wisconsin
51. Wyoming
52. American Samoa
53. Federated States of Micronesia
54. Guam
55. Marshall Islands
56. Northern Mariana Islands
57. Palau
58. Puerto Rico
59. Virgin Islands

<u>BASE: ALL US RESPONDENTS 18+</u>
netRegionUS HIDDEN: Census Region (US Net)

1. Northeast [dmStateUS=7,20,22,30,40,46,31,33,39]
2. Midwest [dmStateUS=14,15,23,36,50,16,17,24,26,28,35,42]
3. South [dmStateUS=8,9,10,11,21,34,41,47,49,1,18,25,43,4,19,37,44]
4. West [dmStateUS=3,6,13,27,29,32,45,51,2,5,12,38,48]
5. Non-US State [dmStateUS=52-59]

<u>BASE: ALL US RESPONDENTS 18+</u>
dmZipUS [Zip Code (US)] What is your zip code?
[5 DIGITS – DISPLAY ERROR IF ZIP CODE IS NOT VALID]

<u>BASE: ALL US RESPONDENTS 18+</u>
dmEduUS [Education (US)] What is the highest level of education you have completed?

[SINGLE RESPONSE]
[PROGRAMMER NOTE: DISPLAY IN ONE COLUMN, GOING DOWN.]

1. Less than high school
2. Completed some high school
3. High school graduate
4. Job-specific training program(s) after high school
5. Some college, but no degree
6. Associate degree
7. Bachelor's degree (such as B.A., B.S.)
8. Some graduate school, but no degree
9. Graduate degree (such as MBA, MS, M.D., Ph.D.)

<u>BASE: ALL US RESPONDENTS 18+</u>
netEduUS Education (US Net)

1. Less than HS degree [dmEduUS/1,2]
2. HS degree to less than 4 year college degree [dmEduUS/3-6]
3. 4 year college degree or more [dmEduUS/7-9]

<u>BASE: ALL US RESPONDENTS 18+</u>
dmHhIncUS [Household Income (US)] How much total combined income did all members of your household earn before taxes last year?

This includes money from jobs; net income from business, farm, or rent; pensions; dividends; interest; social security payments; and any other money income received by members of your household who are eighteen (18) years of age or older.

[PROGRAMMER NOTE: DISPLAY IN ONE COLUMN, GOING DOWN.]

1. Less than $15,000
2. $15,000 to $24,999
3. $25,000 to $34,999
4. $35,000 to $49,999
5. $50,000 to $74,999
6. $75,000 to $99,999

7. $100,000 to $124,999
8. $125,000 to $149,999
9. $150,000 to $199,999
10. $200,000 to $249,999
11. $250,000 or more
99. Prefer not to answer

[PROGRAMMER NOTE: INSERT "Why do we ask this question?" pop-up BELOW CHOICES]

BASE: ALL US RESPONDENTS 18+

netHhIncUS HIDDEN: Household Income (US Net)

1. Less than $15,000
2. $15,000-$24,999
3. $25,000-$34,999
4. $35,000-$49,999
5. $50,000-$74,999
6. $75,000-$99,999
7. $100,000 or more
99. Prefer not to answer

BASE: ALL US RESPONDENTS 18+

dmHispUS [Hispanic Origin (US)] Are you of Hispanic, Latino, or Spanish origin?

1. Yes
2. No
9. Prefer not to answer

[PROGRAMMER NOTE: INSERT "Why do we ask this question?" pop-up BELOW CHOICES]

BASE: ALL US RESPONDENTS 18+

dmRaceUS [Race (US)] What is your race?

[PROGRAMMER NOTE: DISPLAY IN ONE COLUMN.]
[SINGLE RESPONSE]

1. White
2. Black or African American
3. Native American or Alaskan Native
4. South Asian
5. Chinese
6. Korean
7. Japanese
8. Filipino
9. Arab/West Asian
10. Pacific Islander
11. Other Asian
12. Mixed Race
13. Some other race
99. Prefer not to answer

[PROGRAMMER NOTE: INSERT "Why do we ask this question?" pop-up BELOW CHOICES]

BASE: ALL US RESPONDENTS 18+

finRaceUS HIDDEN: Race (US Final)

[IF ANSWERED HISPANIC (dmHispUS/1) ANSWER TO finRaceUS IS CODE14, OTHERWISE finRaceUS=dmRaceUS.]

1. White
2. Black or African American
3. Native American or Alaskan Native
4. South Asian
5. Chinese
6. Korean
7. Japanese
8. Filipino
9. Arab/West Asian
10. Pacific Islander
11. Other Asian
12. Mixed Race
13. Some other race
14. Hispanic
99. Prefer not to answer

BASE: ALL US RESPONDENTS 18+

netRaceUS HIDDEN: Race (US Net)

1. Hispanic [finRaceUS/14]
2. Black (Not Hispanic) [finRaceUS/2]
3. Asian (Not Hispanic) [finRaceUS/4-8,10-11]
4. All Other (Not Hispanic) [finRaceUS/1,3,9,12,13]

BASE: ALL RESPONDENTS AND 18+

dmMarStat [Marital Status] What is your marital status?

1. Never married
2. Married or civil union
3. Divorced
4. Separated
5. Widow/Widower
6. Living with partner

BASE: ALL RESPONDENTS AND 18+

netMarStat HIDDEN: Marital Status (Net)

1. Never married [dmMarStat/1]
2. Married/Living with partner [dmMarStat/2,6]
3. Divorced/Separated/Widowed [dmMarStat/3,4,5]

BASE: ALL US RESPONDENTS 18+

dmAdultHh [Adults in Household]

[IF AGE 18 OR OVER (dmAge/>17)]

Including yourself, how many people age 18 or older live in your household?

|__|__|

BASE: ALL US RESPONDENTS 18+

dmChildHh [Children in Household]

[IF AGE 18 OR OVER (dmAge/>17)]

How many people under the age of 18 live in your household?

|__|__|

BASE: ALL US RESPONDENTS 18+

netHhSize HIDDEN: Size of Household (Net)

1. 1 HH member [sum(dmAdultHh+dmChildHh)=1]
2. 2 HH member [sum(dmAdultHh+dmChildHh)=2]
3. 3 HH member [sum(dmAdultHh+dmChildHh)=3]
4. 4 HH member [sum(dmAdultHh+dmChildHh)=4]
5. 5+ HH member [sum(dmAdultHh+dmChildHh)=5 OR MORE]

BASE: ALL US RESPONDENTS 18+

Q18. Please select your household size including only the people you shop for or with.

[INSERT NUMBER]

BASE: ALL RESPONDENTS 18+

dmEmploy [Employment Status] Which of the following best describes your employment status?

[SINGLE RESPONSE]

1. Employed full time
2. Employed part time
3. Self-employed full time
4. Self-employed part time
5. Not employed, but looking for work
6. Not employed and not looking for work
7. Not employed, unable to work due to a disability or illness
8. Retired
9. Student
10. Stay-at-home spouse or partner

BASE: ALL RESPONDENTS 18+

netEmploy HIDDEN: Employment Status (Net)

1. Employed (FT, PT or Self) [dmEmploy/1,2,3,4]
2. Not Employed [dmEmploy/5-10]

BASE: ALL RESPONDENTS 18+

23. Do you have a car?

1. Yes
2. No