Complete Journey

GROUP 2

Alessandra Vittorini, Marc Poblete, Rebecca Karunakaran, Samuel Peoples

Executive Summary

Given our customer base of roughly 45,000 households, our goal was to segment and profile our customers based on factors such as income, shopping behavior, and demographics; then use this data to identify frequently paired items to develop a Direct Mail Coupon Campaign.

A targeted cluster of 8,500 households was selected as this group used coupons slightly more often than the average customer. This provides the opportunity to persuade customers to purchase items they enjoy more frequently, and thus improve gross revenue.

We identified the most popular combination of items using association rule mining and analyzed the rules using the lift, confidence, support, and conviction metrics.

Given a $10,000 liability margin, our goal was to determine the optimal coupon value that would encourage the consumer base to increase item sales frequency and overall revenue.

Research Objective and Methodology

1. Identifying potential coupon users by segmentation and profiling on customer demographics
2. Measuring the relationship between price, quantity, and coupon value across selected customer group
3. Identifying optimal items sold using apriori rule association across selected customer group
4. Determining optimal coupon values for selected items; meat and food add-ons.
5. Calculating expected gross revenue and lift from coupon promotion.

Tools : K-mean clustering, Association Rule Mining, Python

About the Complete Journey Dataset

Complete Journey Dataset contains household level transaction data for 2 years

household	Identifies unique households
year	Year =1 or Year =2
category	Product Category. Example "baking", "deli"," meat" etc
quantity	Number of the products purchased during the trip
Sales value	Amount of dollars retailer receives from sale
Coupon_discount	Discount applied due to manufacturer coupon
dh_coupon_redeemed	Number of Store coupons redeemed
value of dh coupon	Value of store coupon
Age_Desc	Estimated age range (19-24, 25-34, 35-44,45-54, 55-64, 65+)
Marital_Status_Code	Marital Status (A - Married, B- Single, U - Unknown)
Income_Desc	Household income (100-124K, 125–149K, 150-174K…….)
Homeowner_Desc	Homeowner, renter, Probable Owner, Probable renter, Unknown)
HH_Comp_Dec	Household composition (1 Adult Kids, 2 Adult Kids, Single Female, Single Male, Unknown)
Household_Size_Desc	Size of household up to 5+
Kids_Category_Desc	Number of children present up to 3+

Data set

Data set had mostly Categorical Data which included Product Category and Customer Demographics

Few Numerical Data like Price, Quantity, Coupon Discount

| household_k	year	category	quantity	sales value	COUPON_DISCOUNT	num_dh_coupon_red	value of dh		AGE_DESC	MARITAL	INCOME_	HOMEOWNE	HH_COMP_	HOUSEHOLD	KID_CATEGORY_DESC
22	1	books	1	0		1	0	0	45-54	A	75-99K	Homeowner	2 Adults No K	2	None/Unknown
101	1	other	1	0		1	0	0	45-54	A	Under 15k	Homeowner	2 Adults Kids	4	
208	1	other	1	0		1	0	0	45-54	U	50-74K	Homeowner	2 Adults No K	2	None/Unknown
386	1	health	1	0		0	0	0	35-44	B	25-34K	Probable Ow	Single Fema	1	None/Unknown
1537	1	other	2	0		2	0	0	19-24	A	35-49K	Renter	2 Adults Kids	3	
2162	1	books	1	0		1	0	0	45-54	U	75-99K	Homeowner	Unknown		None/Unknown
2183	1	other	1	0		1	0	0	65+	A	25-34K	Homeowner	2 Adults No K	2	None/Unknown
42	2	water	0	0		0	0	0	65+	U	50-74K	Unknown	Single Male		None/Unknown
57	1	baby	0	0		1	0	0	45-54	B	35-49K	Homeowner	2 Adults No K	2	None/Unknown
139	1	baby	0	0		1	0	0	35-44	B	75-99K	Homeowner	Single Male		None/Unknown
158	2	pet	0	0		1	0	0	45-54	U	50-74K	Homeowner	Unknown		None/Unknown
165	2	none	0	-0.75	-0.75	3	0	0	55-64	A	35-49K	Homeowner	2 Adults No K	2	None/Unknown
214	2	none	0	-7	-7	6	0	0	55-64	B	Under 15k	Homeowner	2 Adults No K	2	None/Unknown
221	1	none	0	-1	-1	2	0	0	65+	U	15-24K	Unknown	Unknown		None/Unknown
281	1	none	0	0		1	0	0	35-44	B	50-74K	Homeowner	Single Fema	1	None/Unknown
290	2	health	0	0		1	0	0	25-34	U	35-49K	Unknown	2 Adults No K	2	None/Unknown
346	1	none	0	-2	-2	2	0	0	65+	A	50-74K	Homeowner	2 Adults No K	2	None/Unknown
346	1	frozen_piz	0	0		1	0	0	65+	A	50-74K	Homeowner	2 Adults No K	2	None/Unknown
426	2	none	0	-7.25	-7.25	13	0	0	45-54	U	50-74K	Homeowner	2 Adults No K	2	None/Unknown

Data Transformation

The categorical columns in the original data were transformed using one-hot encoding, which transforms a categorical attribute with *n* labels into a vector of size *n-1* with only one of the values as 1 (active). The lowest-quantity column is dropped to avoid the dummy-variable trap.

K- Number of Clusters

The Elbow Method for K means clustering was used to determine an optimal number of clusters of 3
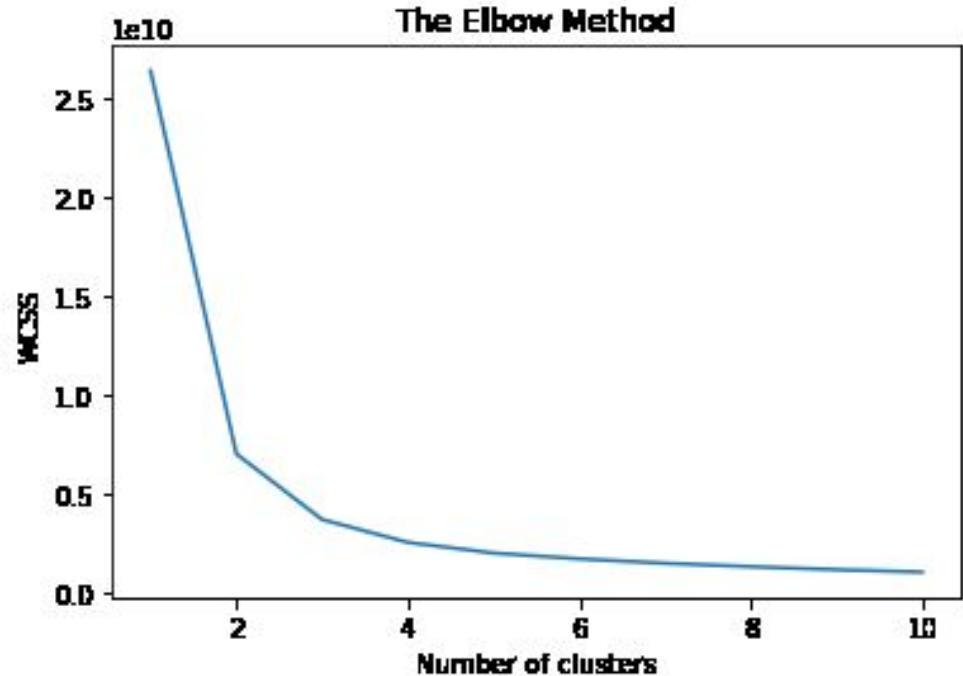
Customer Segments and Profile

|		
Buys everything	Buys products for baby and children, and fresh food	Buys fresh food and household items mainly
Least likely to use coupons	Most likely to use coupons	Marginally more likely to use coupons
Age distribution general	Age Group 35-54, few 19-24 and 25-34	Age Group 35-54
Marital Status general	Mostly Married with large families	Mostly Married

Data Findings

- Focused on **Cluster/Segment 2**: - Average Coupon Discount of $3.94
  - Slightly above average coupon rate of -$3.52
  - Cluster 2 has 8,329 households of 46,757 total

Average of COUPON_DISCOUNT	
Cluster ▾	Total
0	-2.605079365
1	-7.095120226
2	-3.943273114
Grand Total	-3.516840656

Data Findings
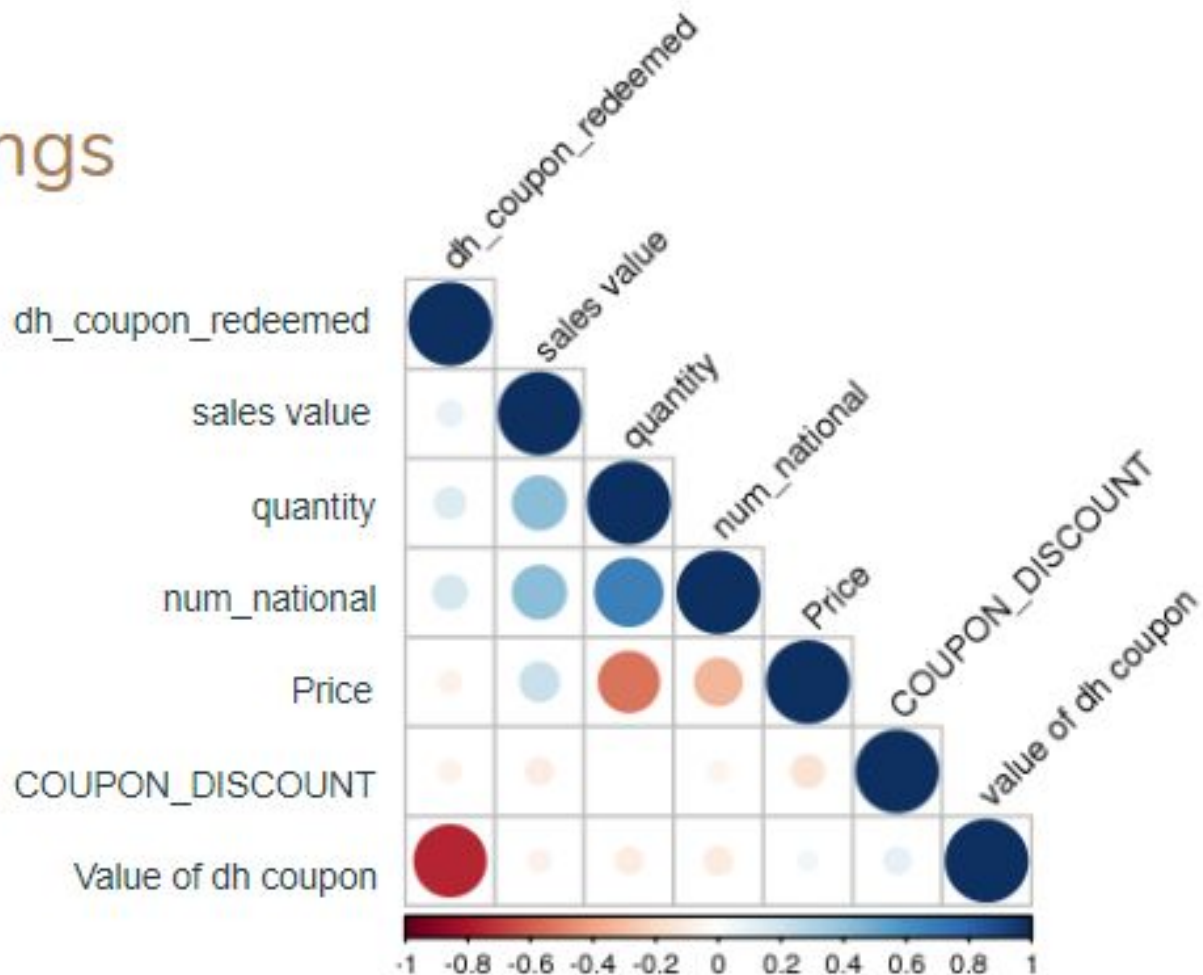
Data Findings

- <u>Performed Apriori Rule Association</u>
  - **Confidence** (conf): how often the rule has been found to be true
  - **Support** (supp): how frequently the itemset appears in the dataset
  - **Lift** (lift): ratio of observed support to expected if x & y were independent
  - **Conviction** (conv): ratio of expected frequency that x occurs without y
- <u>Segment (Cluster) 2:</u>
  - drinks, frozen_pizza   ->   meat
  - baking, food   ->   food_add-ons
  - dessert, packaged_foods ->   meat

Cluster	Cart1 ->	Cart2	conf	supp	lift	conv
2	drinks, frozen_pizza	meat	0.943	0.085	1.265	4.454
2	baking, food	food_add-ons	0.928	0.217	1.417	4.803
2	dessert, packaged_foods	meat	0.928	0.083	1.244	3.512

Calculating Coupon Value

- Expected participants: 8500
- Maximum liability: $10,000
- Average price of meat: $3.58
  - Average coupon value: $0.26
  - Discount: **7.3%**
  - Num offers: 2
- Average price of food add-on: $1.97
  - Average coupon value: $0.28
  - Discount: **14.0%**
  - Num offers: 1

$Liability =$
8500
$*((2*3.58*MeatDiscount)$
$+(1*1.97*FoodAdd\text{-}OnDiscount))$

Calculating Coupon Value

Food Add-On Discount optimized at 7%, with a meat discount of 14% using a sensitivity matrix.

meat/food	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07
0	0	166.6	333.2	499.8	666.4	833	999.6	1166.2
0.01	608.6	775.2	941.8	1108.4	1275	1441.6	1608.2	1774.8
0.02	1217.2	1383.8	1550.4	1717	1883.6	2050.2	2216.8	2383.4
0.03	1825.8	1992.4	2159	2325.6	2492.2	2658.8	2825.4	2992
0.04	2434.4	2601	2767.6	2934.2	3100.8	3267.4	3434	3600.6
0.05	3043	3209.6	3376.2	3542.8	3709.4	3876	4042.6	4209.2
0.06	3651.6	3818.2	3984.8	4151.4	4318	4484.6	4651.2	4817.8
0.07	4260.2	4426.8	4593.4	4760	4926.6	5093.2	5259.8	5426.4
0.08	4868.8	5035.4	5202	5368.6	5535.2	5701.8	5868.4	6035
0.09	5477.4	5644	5810.6	5977.2	6143.8	6310.4	6477	6643.6
0.1	6086	6252.6	6419.2	6585.8	6752.4	6919	7085.6	7252.2
0.11	6694.6	6861.2	7027.8	7194.4	7361	7527.6	7694.2	7860.8
0.12	7303.2	7469.8	7636.4	7803	7969.6	8136.2	8302.8	8469.4
0.13	7911.8	8078.4	8245	8411.6	8578.2	8744.8	8911.4	9078
0.14	8520.4	8687	8853.6	9020.2	9186.8	9353.4	9520	9686.6

Calculation of Expected Lift

- Average yearly sales for drinks, frozen pizza, baking, food, dessert, packaged foods, meat, and food add-ons yielded $1,062.34 in gross revenue.
- In an ideal scenario with 100% participation, sales for this item averaged across the year with the modified discounts yields an estimated gross revenue of $6,803.23
- This represents an ideal lift of $5,740.90, a **137%** increase in sales.

Recommendation

- Customers identified in Cluster 2 are the target group for this promotion.
  - These customers tend to have smaller families, are married between 35 and 54, and use coupons slightly more often than the average customer.
- If customers buy:
  - a drink and frozen pizza, meat is 14% off.
  - a baking item and any food item, add-ons are 7% off.
  - a dessert and a packaged food, meat is 14% off.
- Promotions can be combined for redemption against one item for each coupon.
- Run campaign for one month to determine if expected liability is met, and targeted customer group is responding to the promotion.

References

- Data source: https://www.dunnhumby.com/careers/engineering/sourcefiles
- SKLearn: https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html
- Efficient Apriori: https://pypi.org/project/efficient-apriori/