Sam's Club First-Year Renewal Analysis Report

DSO 545 Statistical Computing and Data Visualization Instructor: Abbass Sharif

Mengtian (Monica) Hu

November 2018

Catalogue

1. INTRODUCTION	3
• Problem	3
• Use of data	3
 Deliverables 	3
 Summary 	3
2. LOGISTIC REGRESSSION MODEL	4
 Input variables 	4
Behavior variables	4
Demographic variables	4
• Procedure steps	4
3. DECISION TREE	5
4. CLUSTER EXPLORATION	6
5. INTERESTING FINDINGS	9
6. CUSTOMER PROFILES	
 Lazy buyers 	12
 Golden buyers 	13
 Active buyers 	14
 Non-target buyers 	14
7. APPENDIX – SHINY DASHBOARD	16

1. INTRODUCTION

> Problems

Based on historical customer behaviors and demographic attributes, which behaviors drive first-year members to renew, and what does an "ideal" first-year member look like?

➤ Use of data

- 1. Member demographic data since Jan. 2017
- 2. Point of Sale (POS) data for those members, from join date to DTR date; item level data for every transaction
- 3. Item/category cross-reference data

> Deliverables

- 1. A shiny dashboard to visualize our clusters
- 2. An EDA (exploratory data analysis) model
- 3. A PowerPoint presentation with business insights and recommendations

> Summary

We've decided to use logistic regression to find expert variables first. Using these variables, we've identified 3 customer segments with the highest possibility of renewal using regression and decision tree. Among the three segments, we analyzed customer demographic information along with their purchase behaviors. Based on our findings, we've put together an ideal customer profile and provided business suggestions. Our report includes 5 parts:

- 1. Logistic regression
- 2. Decision tree customer segmentation
- 3. Cluster exploration
- 4. Interesting findings
- 5. Customer profile descriptions
- 6. Appendix R Shiny Dashboards

2. LOGISTIC REGRESSION MODEL

The goal of performing a regression is to find significant variables that can drive FY customer to renew one year later. We treat members as our observations and their attributes as our input variables. The model we used is machine learning logistic regression. Our outcome Y is 0 and 1, where 0 represents not renew and 1 represents renew, including both basic and plus renew.

Input variables

> Behavior Variables

Sales (member purchase of the whole year), Total Number of Visit, Total Number of Product Purchased, Total Number of Category, Category Number, Auto-Renew, Days since Last Visit (days between last visit and Jan-31 2018), Tender Type Sales

> Demographic Variables

Income, Household Size, Ethnic, Plus Membership, Payroll Deduct, Marital Status, Miles to Sam's Club, Membership Type.

Procedure steps

- 1. After grouping POS data by membership ID and product category, the regression result showed that product category was not a significant variable. We then removed product category number from our input variables.
- 2. Added tender type sales amount to the logistic regression model.
- 3. Eliminated the collinearity influence among variables.
- 4. The table below was the result of the regression performed—significant variables and their coefficients. A positive coefficient means that the variable has a positive influence on the renewal rate, and a negative coefficient means that the variable has a negative influence on the renewal rate.

Name	Coefficient
Total Number of Product Purchase	11.550083
Auto Renew	5.21593252
Payroll Deduct	4.93031114
Tender - Coupon	2.38825175
Tender - Debit Card	1.86699783
Sales	1.74374754
Days since Last Visit	-2.4506091
Tender - EBT Food stamps	-6.7810785

Since we had 280-thousand records, we separated the data into two parts: 50% used in model training and 50% in testing. The testing process showed that the model had a 77% accuracy rate, which was significant enough for us to do the following steps.

3. DECISION TREE

The goal of our decision tree model was to segment customers according to their renewal status. The impurity of renew status should be the largest among segments and the smallest within each segment.

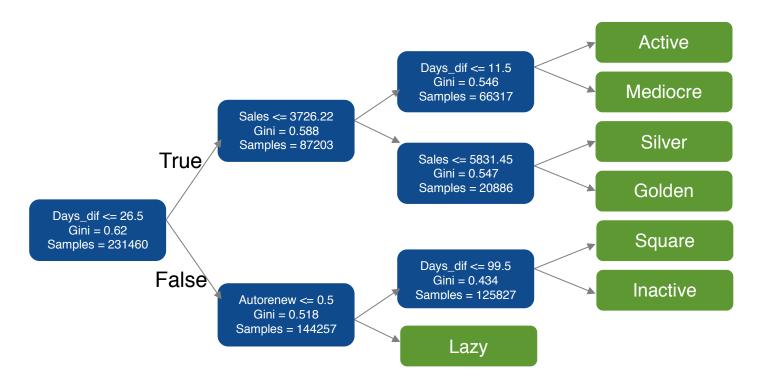
The model we select was machine learning decision tree model using Python. The method we used in segmentation is Gini coefficient.

In this model, our objective Y is renewed or not (1 and 0), our input X are significant variables from step 2, the regression model.

The decision tree model automatically used the most significant 3 variables to make the segmentations:

- 1. Sales
- 2. Auto-Renew
- 3. Days since the last visit

We have the following 7 segments as results from the decision tree:



Segment description:

Segments	Auto-Renew	Purchase	Days Since Last Visit
Lazy Buyers	Yes		Above 27
Golden Buyers		Above \$5,800	Below 27
Active Buyers		Below \$3,700	Below 12
Silver Buyers		Btw \$3,700 and \$5,800	Below 27
Mediocre Buyers		Below \$3,700	Btw 12 and 27
Square Buyers	No		Btw 27 and 100
Inactive Buyers	No		Above 100

The renew rate of seven segments is shown in the table below. The top three segments have the highest renew rate, so customers in these three segments are our ideal customers.

Customer Clusters	Renew Rate
Lazy Buyers	0.986477
Golden Buyers	0.941133
Active Buyers	0.912442
Silver Buyers	0.898198
Mediocre Buyers	0.691307
Square Buyers	0.413817
Inactive Buyers	0.188465

4. CLUSTER EXPLORATION

Now, we have already used logistic regression and decision tree to effectively segment customers into 7 groups. Based on their renewal rates and characteristics, we gave them nicknames accordingly:

- ➤ Lazy Buyers auto-renew
- ➤ Golden Buyers actively renew (ranked 1)
- ➤ Active Buyers actively renew (ranked 2)
- ➤ Silver Buyers actively renew (ranked 3)
- ➤ Mediocre Buyers general renew (rank 1)
- ➤ Square Buyers general renew (rank 2)
- ➤ Inactive Buyers inactively renew

Besides the three significant variables we used, there were many other useful variables such as product categories, tender type and age, income (demographic data). Next, we used these data to identify characteristics for each of the seven customer segments using statistical methods:

> First, for each variable, we calculated the average value of each customer segment and that of all Sam's club customers:

(19 product categories in Sam's Club)

CATEGORIES	Lazy Buyers	Golden Buyers	Active Buyers	Silver Buyers	Mediocre Buyers	Square Buyers	Inactive Buyers	All
DMM	60.54	1404.94	52.94	72.68	52.01	54.54	56.16	210.09
DMM ADULT BEVERAGES	82.51	440.89	94.07	170.33	83.67	80.26	79.90	122.11
DMM APPAREL AND JEWELRY	107.30	428.44	112.76	215.44	105.78	109.57	74.32	135.57
DMM CANDY-SNACKS-BEV-TOB-WHSL	176.21	2252.65	238.68	585.03	220.62	223.61	136.50	320.31
DMM CONSUMABLES	172.35	1108.95	230.37	458.26	208.61	188.95	97.67	246.99
DMM CONSUMER ELECTRONICS	281.07	606.66	182.92	339.64	196.99	259.71	283.05	286.73
DMM DRY GROCERY	126.82	1087.34	166.42	383.67	159.55	157.03	90.98	205.01
DMM ENTERTAINMENT AND OFFICE S	99.52	530.26	73.72	131.89	74.48	115.78	145.71	140.41
DMM FREEZER - DELI - COOLER	127.87	1053.12	183.01	465.15	169.32	161.61	84.61	214.04
DMM FUEL	136.75	936.53	382.78	671.18	271.65	190.70	86.89	330.91
DMM HBA - OTC AND BABY	130.81	574.56	180.52	347.08	162.04	142.26	86.56	183.70
DMM HEALTH CARE	434.81	539.32	130.94	266.98	119.76	186.60	301.17	242.00
DMM HOME - HARDLINES - TIRE -	189.55	583.97	170.73	332.46	158.43	169.00	139.38	208.69
DMM MEAT - SEAFOOD	130.11	1025.29	168.51	385.60	162.68	170.12	99.29	215.61
DMM MOBILE AND SERVICES	296.61	635.75	155.22	423.18	184.00	295.31	363.68	316.67
DMM PREPARED MEALS	22.34	259.60	32.43	54.87	27.56	25.32	22.39	43.94
DMM PRODUCE AND BAKERY	78.78	698.73	133.29	325.50	113.79	101.95	47.44	140.98
DMM SEASONAL	96.66	278.04	84.75	156.78	83.68	89.23	73.85	108.61
DMM TOBACCO	663.12	35158.34	229.17	461.38	167.58	1875.12	5457.53	8429.04

(Demographic data)

Demographic Data	Lazy Buyers	Golden Buyers	Active Buyers	Silver Buyers	Mediocre Buyers	Square Buyers	Inactive Buyers	All
Miles away avg	56.03	19.81	27.14	16.27	27.66	33.04	64.53	41.40
Age avg	49.32	45.71	48.87	46.34	48.29	47.23	47.61	47.82
Marital rate avg	0.85	0.89	0.84	0.86	0.83	0.82	0.79	0.82
Income lower interval avg	67.49	75.02	67.05	72.93	66.01	64.46	63.99	66.10
Income upper interval avg	73.02	77.69	72.69	76.74	71.36	69.97	68.04	70.92
Household size avg	3.00	3.17	2.99	3.12	2.96	2.96	2.88	2.96
Nbr of children avg	0.47	0.58	0.48	0.57	0.48	0.49	0.45	0.48
Plus membership rate avg	0.27	0.80	0.32	0.59	0.32	0.32	0.27	0.33
Auto renew avg	1.00	0.28	0.29	0.28	0.20	0.00	0.00	0.18
Payroll deduct avg	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.01

(Tender type)

Tender Type	Lazy Buyers	Golden Buyers	Active Buyers	Silver Buyers	Mediocre Buyers	Square Buyers	Inactive Buyers	All
American Express	3078.35	41008.89	2292.79	5105.52	2263.43	2717.23	2645.24	5618.68
Associate Discount Card	8.48	45.74	12.67	20.32	10.06	15.13	24.94	19.20
Cash (change)	3676.72	73239.32	5087.50	13055.27	4860.35	5691.61	4074.36	8568.50
Check	6684.55	16760.26	1056.04	981.62	1603.91	1452.87	2538.61	3958.49
Coupon	537.45	796.12	418.47	547.10	435.82	469.01	482.03	501.11
DOTCOM	3473.14	10763.69	1305.87	2366.55	1394.67	1584.65	2224.65	3027.28
Debit Card	11266.86	115052.57	15178.36	39504.74	13637.94	13776.85	7134.97	18320.16
Discover	2145.85	15262.48	1792.27	3855.26	1811.39	2172.78	1894.99	3093.68
EBT Cash	697.60	995.60	559.47	1000.23	662.28	1047.96	959.10	855.38
EBT Foodstamps	2985.71	15825.21	2981.33	9304.16	4238.70	5556.79	4081.70	5378.18
Electronic Check (ECA)	4011.63	70036.50	2783.19	4640.68	3059.61	6084.75	8109.50	11765.07
Fleet	0.00	164.56	121.78	0.00	29.55	29.00	0.00	120.71
Mastercard	3680.97	46216.28	3376.51	8482.37	3272.76	3342.42	2650.56	6084.11
Payroll Check	466.42	852.16	350.41	634.34	401.70	576.91	306.04	530.12
Rx third party sales	988.99	1922.13	970.17	1196.63	865.60	936.69	893.90	1126.35
Sams Business Credit	11981.02	90918.66	2553.70	3927.56	3087.17	6957.64	17280.26	19519.14
Sams Consumer Credit	5344.89	33760.05	2816.93	6414.69	2577.14	2540.23	2737.82	5176.86
Sams Direct Credit	2774.74	15335.90	1194.00	1554.74	1189.13	1487.95	1678.35	3918.05
Service Income	1092.99	1378.63	807.20	899.29	822.97	973.69	1178.44	995.72
Shopping Card	656.72	4848.75	684.73	1584.33	677.90	728.37	478.78	948.63
Unknown	4779.05	6533.69	3862.65	4341.93	3846.90	4493.68	5380.70	4675.95
Visa	6417.75	94577.99	8159.69	21617.86	7252.07	7014.63	4015.31	11753.26
Wal-Mart Credit Card	1577.71	2676.79	825.61	1160.90	952.83	1119.78	1250.01	1244.34
Coin Voucher	0.00	0.00	0.00	0.00	69.83	227.78	0.00	148.80

➤ Second, for each variable, we scaled averages of the seven customer segments to the percentage of that of all customers' average, and then minus 1 to visualize the difference. Next, we selected segments with the most significant difference compare to other segments and used **red** to represent "significantly lower than the average" and **green** to represent "significantly higher than the average":

(19 product categories in Sam's Club)

CATEGORIES	Lazy Buyers	Golden Buyers	Active Buyers	Silver Buyers	Mediocre Buyers	Square Buyers	Inactive Buyers	All
DMM	-0.71	5.69	-0.75	-0.65	-0.75	-0.74	-0.73	1
DMM ADULT BEVERAGES	-0.32	2.61	-0.23	0.39	-0.31	-0.34	-0.35	1
DMM APPAREL AND JEWELRY	-0.21	2.16	-0.17	0.59	-0.22	-0.19	-0.45	1
DMM CANDY-SNACKS-BEV-TOB-WHSL	-0.45	6.03	-0.25	0.83	-0.31	-0.30	-0.57	1
DMM CONSUMABLES	-0.30	3.49	-0.07	0.86	-0.16	-0.23	-0.60	1
DMM CONSUMER ELECTRONICS	-0.02	1.12	-0.36	0.18	-0.31	-0.09	-0.01	1
DMM DRY GROCERY	-0.38	4.30	-0.19	0.87	-0.22	-0.23	-0.56	1
DMM ENTERTAINMENT AND OFFICE S	-0.29	2.78	-0.48	-0.06	-0.47	-0.18	0.04	1
DMM FREEZER - DELI - COOLER	-0.40	3.92	-0.14	1.17	-0.21	-0.24	-0.60	1
DMM FUEL	-0.59	1.83	0.16	1.03	-0.18	-0.42	-0.74	1
DMM HBA - OTC AND BABY	-0.29	2.13	-0.02	0.89	-0.12	-0.23	-0.53	1
DMM HEALTH CARE	0.80	1.23	-0.46	0.10	-0.51	-0.23	0.24	1
DMM HOME - HARDLINES - TIRE -	-0.09	1.80	-0.18	0.59	-0.24	-0.19	-0.33	1
DMM MEAT - SEAFOOD	-0.40	3.76	-0.22	0.79	-0.25	-0.21	-0.54	1
DMM MOBILE AND SERVICES	-0.06	1.01	-0.51	0.34	-0.42	-0.07	0.15	1
DMM PREPARED MEALS	-0.49	4.91	-0.26	0.25	-0.37	-0.42	-0.49	1
DMM PRODUCE AND BAKERY	-0.44	3.96	-0.05	1.31	-0.19	-0.28	-0.66	1
DMM SEASONAL	-0.11	1.56	-0.22	0.44	-0.23	-0.18	-0.32	1
DMM TOBACCO	-0.92	3.17	-0.97	-0.95	-0.98	-0.78	-0.35	1

(Demographic data)

(2 cm sgrapm c and)											
Demographic Data	Lazy Buyers	Golden Buyers	Active Buyers	Silver Buyers	Mediocre Buyers	Square Buyers	Inactive Buyers	All			
Miles away avg	0.35	-0.52	-0.34	-0.61	-0.33	-0.20	0.56	1			
Age avg	0.03	-0.04	0.02	-0.03	0.01	-0.01	0.00	1			
Marital rate avg	0.04	0.08	0.02	0.05	0.01	-0.01	-0.03	1			
Income lower interval avg	0.02	0.13	0.01	0.10	0.00	-0.02	-0.03	1			
Income upper interval avg	0.03	0.10	0.02	0.08	0.01	-0.01	-0.04	1			
Household size avg	0.01	0.07	0.01	0.05	0.00	0.00	-0.03	1			
Nbr of children avg	-0.02	0.21	-0.01	0.20	-0.01	0.01	-0.07	1			
Plus membership rate avg	-0.20	1.41	-0.04	0.76	-0.05	-0.04	-0.18	1			
Auto renew avg	4.67	0.59	0.64	0.61	0.13	-1.00	-1.00	1			
Payroll deduct avg	-1.00	0.36	0.67	0.17	-0.23	-0.36	0.21	1			

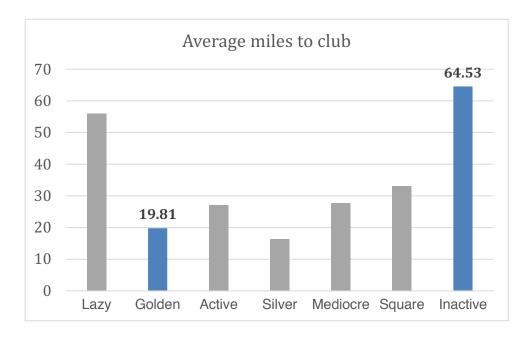
(Tender type)

DSO 545 Final Project – Sam's Club First Year Renewal Analysis Report

Tender Type	Lazy Buyers	Golden Buyers	Active Buyers	Silver Buyers	Mediocre Buyers	Square Buyers	Inactive Buyers	All
American Express	-0.45	6.30	-0.59	-0.09	-0.60	-0.52	-0.53	1
Associate Discount Card	-0.56	1.38	-0.34	0.06	-0.48	-0.21	0.30	1
Cash (change)	-0.57	7.55	-0.41	0.52	-0.43	-0.34	-0.52	1
Check	0.69	3.23	-0.73	-0.75	-0.59	-0.63	-0.36	1
Coupon	0.07	0.59	-0.16	0.09	-0.13	-0.06	-0.04	1
DOTCOM	0.15	2.56	-0.57	-0.22	-0.54	-0.48	-0.27	1
Debit Card	-0.39	5.28	-0.17	1.16	-0.26	-0.25	-0.61	1
Discover	-0.31	3.93	-0.42	0.25	-0.41	-0.30	-0.39	1
EBT Cash	-0.18	0.16	-0.35	0.17	-0.23	0.23	0.12	1
EBT Foodstamps	-0.44	1.94	-0.45	0.73	-0.21	0.03	-0.24	1
Electronic Check (ECA)	-0.66	4.95	-0.76	-0.61	-0.74	-0.48	-0.31	1
Fleet	-1.00	0.36	0.01	-1.00	-0.76	-0.76	-1.00	1
Mastercard	-0.39	6.60	-0.45	0.39	-0.46	-0.45	-0.56	1
Payroll Check	-0.12	0.61	-0.34	0.20	-0.24	0.09	-0.42	1
Rx third party sales	-0.12	0.71	-0.14	0.06	-0.23	-0.17	-0.21	1
Sams Business Credit	-0.39	3.66	-0.87	-0.80	-0.84	-0.64	-0.11	1
Sams Consumer Credit	0.03	5.52	-0.46	0.24	-0.50	-0.51	-0.47	1
Sams Direct Credit	-0.29	2.91	-0.70	-0.60	-0.70	-0.62	-0.57	1
Service Income	0.10	0.38	-0.19	-0.10	-0.17	-0.02	0.18	1
Shopping Card	-0.31	4.11	-0.28	0.67	-0.29	-0.23	-0.50	1
Unknown	0.02	0.40	-0.17	-0.07	-0.18	-0.04	0.15	1
Visa	-0.45	7.05	-0.31	0.84	-0.38	-0.40	-0.66	1
Wal-Mart Credit Card	0.27	1.15	-0.34	-0.07	-0.23	-0.10	0.00	1
Coin Voucher	-1.00	-1.00	-1.00	-1.00	-0.53	0.53	-1.00	1

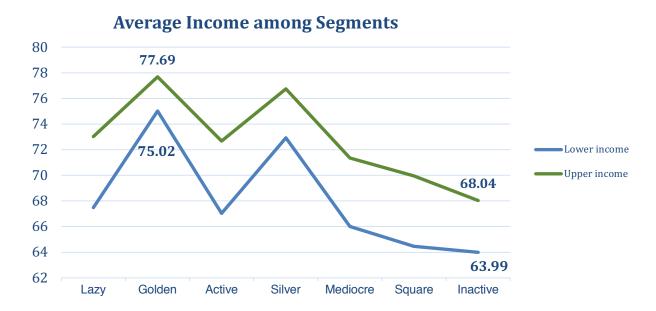
5. INTERESTING FINDINGS

➤ Inactive buyers tend to live farther than the other segments. Golden buyers live relatively nearer than other segments.

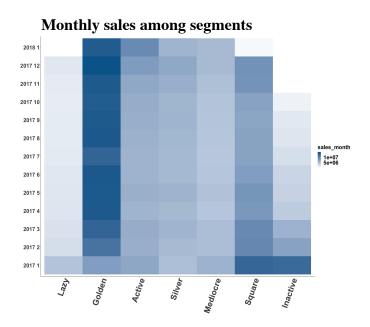


> When it comes to income, Golden Buyers have the highest average income among

these segments, and Inactive Buyers have the least average income.

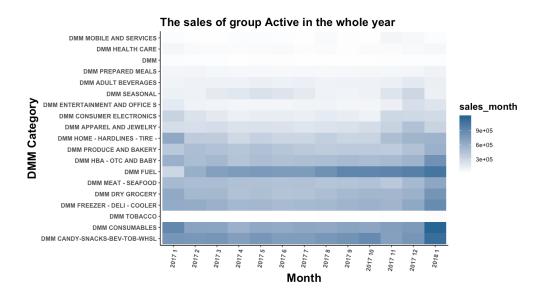


➤ The Inactive Buyers did little shopping in Nov. and Dec. Active Buyers shop the most in Jan. 2018. The Golden Buyers had an increasing shopping trend in the whole year, Inactive Buyers had a decreasing shopping trend in the whole year. The related recommendation is that Sam's Club can do intervene twice during the year, the first is 10 months later after a member join club and the second is 12 months later.



Active buyers tend to buy more and more fuel during the whole year and they shop

more before the renewal deadline



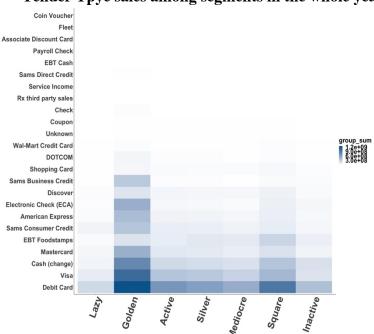
Tobacco has the highest sales in the whole year. Golden buyers buy a large amount of tobacco. Surprisingly, so do inactive buyers. However, inactive buyers only do in the first three months of their membership and then we see a sudden decrease in tobacco purchase. We concluded that tobacco might be one of the reasons that they start the membership, but they don't want to continue it.

DMM sales among segments whole year

DMM sales among segments first 3 mon



Members who used electronic checks, Sam's consumer credits, Sam's business credits, electronic check, American Express, and Mastercard are our ideal customers (Golden Buyers). What's more, most EBT Food stamps users are Square Buyers, which has a low renewal rate within its segment.



Tender Tpye sales among segments in the whole year

6. CUSTOMER PROFILES

Based on our findings, we created the following ideal customer profiles:

Lazy Buyers: Renewal rate – 99%

- ➤ Income range: \$68K~\$73K
- ➤ Household distance: 56 miles
- > Days since the last visit: above 27
- > Auto-renew
- > The 1st month after joining: large amount purchase
- After the 1st month: no difference between the monthly purchase
- ➤ Demographic characteristics:
 - Payroll deduct significantly lower than that of all customers' average Auto-renew rate significantly higher than that of all customers' average
- 19 product categories characteristics:
 DMM purchase significantly lower than that of all customers' average
 DMM FUEL significantly lower than that of all customers' average
 DMM TOBACCO significantly lower than that of all customers' average
 DMM HEALTH CARE significantly higher than that of all customers' average

> Tender type prefence:

The usage of Associate Discount Card, Cash (change), Electronic Check (ECA) and Fleet significantly lower than that of all customers' average

The usage of Check significantly higher than that of all customers' average

Golden Buyers: Renew rate – 94%

➤ Income range: \$75K~\$78K

➤ Household distance: about 20 miles

> Days since the last visit: below 27

Purchase in the whole year: above \$5,800

- > Tobacco lover: significantly higher tobacco purchase compared to other segments
- > Increasing purchase trend
- ➤ Tender type: American Express, Mastercard, electronic check, Sam's credit card
- > Demographic characteristics:

Miles away (distance) significantly lower than that of all customers' average

Auto-renew rate significantly higher than that of all customers' average

Plus membership register rate significantly higher than that of all customers' average

➤ 19 product categories characteristic:

All 19 DMM product categories' purchases significantly higher than that of all customers' average

➤ Tender type prefence:

The usage of American Express, Associate Discount Card, Cash (change), Check, Coupon, DOTCOM, Debit Card, Discover, EBT Foodstamps, Electronic Check (ECA), Mastercard, Payroll Check, Rx third party sales, Sams Business Credit, Sams Consumer Credit, Sams Direct Credit, Shopping Card, Visa and Wal-Mart Credit Card significantly higher than that of all customers' average

Active Buyers: Renew rate – 91%

➤ Income range: \$67K~\$73K

➤ Household distance: about 27 miles

> Days since the last visit: below 12

➤ Purchase: below \$3,700

- Fuel lover significantly higher fuel purchase compared to other segments
- > Increasing purchase trend in fuel
- > Frequent shopping before the renew deadline
- > Demographic characteristics:

Payroll deduct significantly higher than that of all customers' average Auto-renew rate significantly higher than that of all customers' average

➤ 19 product categories characteristic:

DMM purchase significantly lower than that of all customers' average

DMM MOBILE AND SERVICES significantly lower than that of all customers' average

DMM TOBACCO significantly lower than that of all customers' average

> Tender type prefence:

The usage of American Express, Associate Discount Card, Check, DOTCOM, Electronic Check (ECA), Sams Business Credit and Sams Direct Credit significantly lower than that of all customers' average

Following is our non-target customer profile:

Non-target Buyers: Renew rate – 19%

- ➤ Income range: \$64K~\$68K
- ➤ Household distance: about 65 miles
- ➤ Not auto-renew
- ➤ Days since the last visit: around 100
- ➤ High tobacco purchase in the first three months then decrease dramatically
- > Decreasing purchase trend
- > Few transaction before the renew deadline
- Use food stamps
- > Demographic characteristics:

Miles away (distance) significantly higher than that of all customers' average Auto-renew rate significantly lower than that of all customers' average

➤ 19 product categories characteristic:

DMM purchase significantly lower than that of all customers' average

DMM CANDY-SNACKS-BEV-TOB-WHSL significantly lower than that of all customers' average

DMM CONSUMABLES significantly lower than that of all customers' average

DMM DRY GROCERY significantly lower than that of all customers' average

DMM FUEL significantly lower than that of all customers' average

DMM HBA - OTC AND BABY significantly lower than that of all customers' average

DMM MEAT - SEAFOOD significantly lower than that of all customers' average

DMM PRODUCE AND BAKERY significantly lower than that of all customers'

average

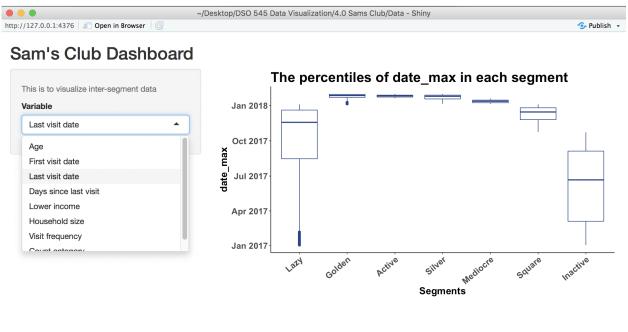
DMM TOBACCO significantly lower than that of all customers' average

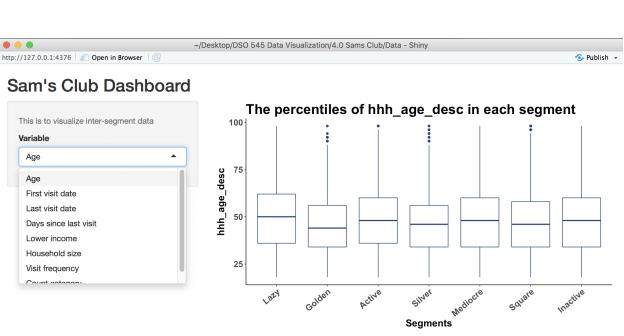
> Tender type prefence:

The usage of American Express, Cash (change), Debit Card, Fleet, Mastercard, Sams Direct Credit, Shopping Card and Visa significantly lower than that of all customers' average

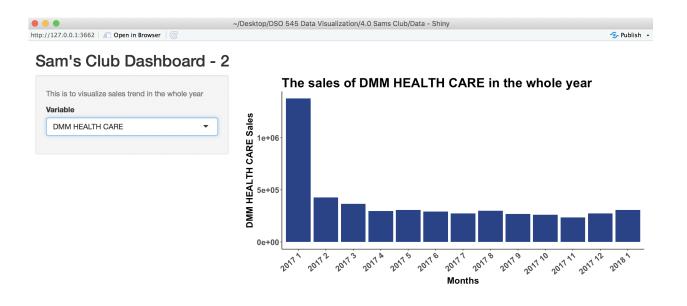
7. APPENDIX - SHINY DASHBOARD

1. The first dashboard is to show the distribution of each demographic variable among our 7 segments.





2. The second dashboard is to show the monthly sales trend when you select a certain DMM product category.



3. The third dashboard is to show the monthly sales trend in all DMM categories when you select a certain segment.

