Marketing Predictive Analytics for a Leading Pizza Chain of the U.S.

By

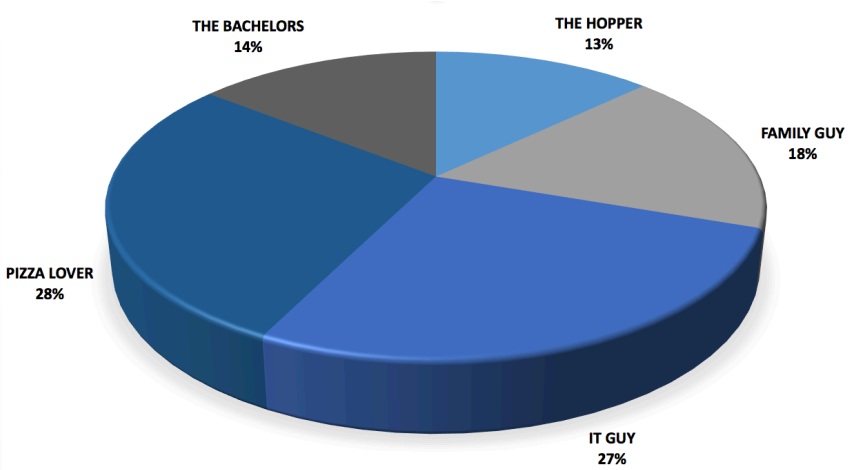
Surendra Potupureddy

Objective 1

Define bucket groups for store profiles and determine offer sets most likely to work for each store profile group

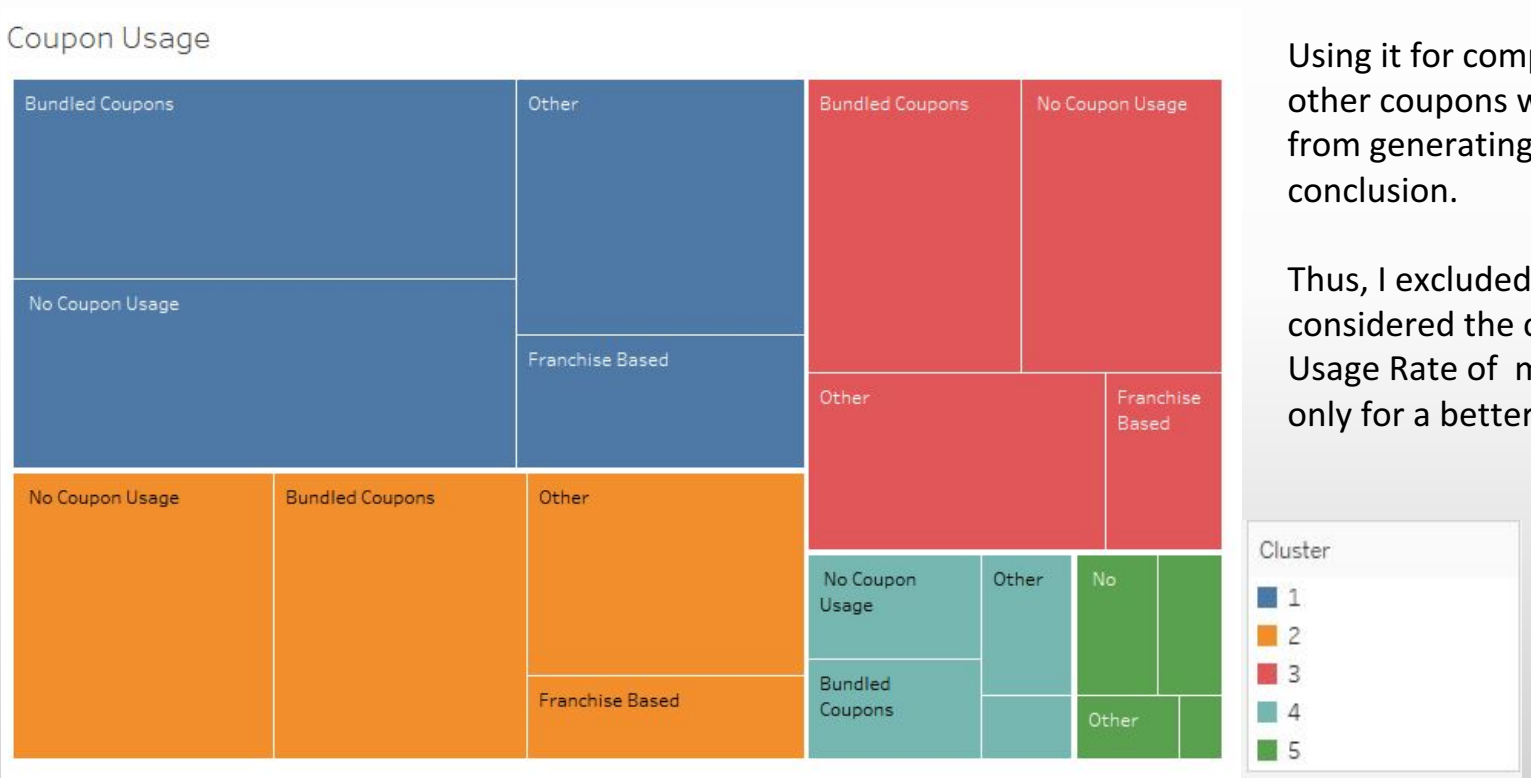
- My goal is to better understand which offers should be promoted in each store profile group
- Segmented stores based on their demographic data and analyzed transactional data to determine coupon sets that best fits each group
- Data set used: January'16 – December'16 food chain data containing both demographic data and transactional data pertaining to each purchase
- The idea for this is that when we get a new store's data it will automatically be labeled with a store profile group
- Aim is to suggest offer sets that have worked best in other stores that have a similar profile

699 distinct stores
5 store profiles
Created based on Customer Purchase Pattern



CLUSTERS	NUMBER OF STORES	PURCHASE POWER INCOME	ORDER AMOUNT	INTERNET USAGE	NUMBER OF CHILDREN	CHURN RATE
The Hopper	90	Moderate	Very High	High	High	Very High
Family Man	122	Very High	High	Low	Very High	High
IT guy	189	Very Low	Moderate	Very High	Moderate	Moderate
Pizza Lover	198	Low	Very Low	Moderate	Very Low	Low
The Bachelors	100	High	Low	Very Low	Low	Very Low

258,212 Transactions
68% Transactions done using a coupon
2000+ Distinct Coupons



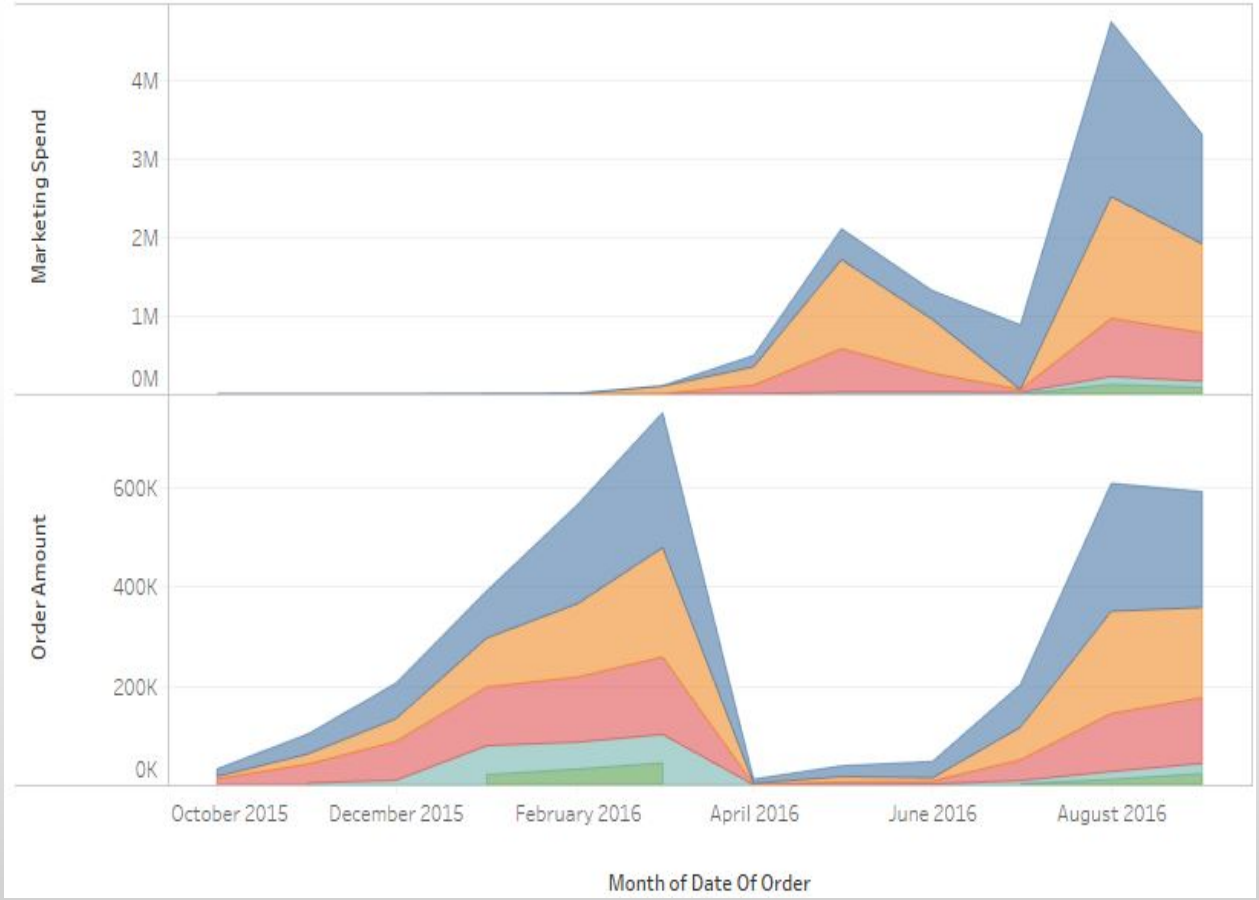
Analysing the chart, **Coupon No:9193** was found responsible for approx. **30%** of the Total Revenue.

Using it for comparison with other coupons will handicap us from generating a logical conclusion.

Thus, I excluded the coupon and considered the coupons with a Usage Rate of more than 1.7% only for a better Analysis

Marketing Cost vs Sales

Across A Period For Each Cluster



It is observed that Marketing Spend has huge impact on the total sales generated across a time period which can be observed in the coming 3 months

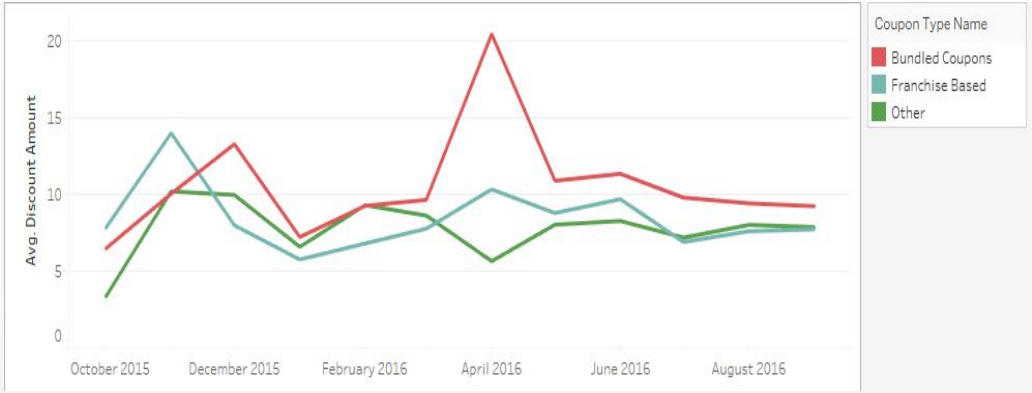
Reducing the expenditure to nearly zero for the months November'15 to March'16 is affecting the sales from April'16 to July'16

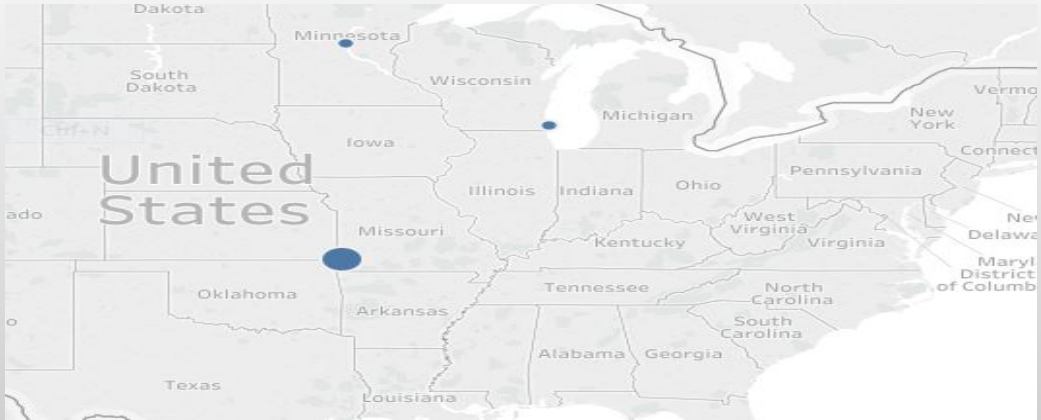
Increase in spend for the months March'16 to July'16 is creating a huge impact in sales growth from July'16 to September'16.

	Oct-15	Nov-15	Dec-15	Jan-16	Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16	Aug-16	Sep-16
Rejuvenated OL	58	122	325	638	1,386	1,933	26	47	86	638	2,086	1,704
Frequent OL	6	17	24	33	40	49	198	234	269	319	688	846
Total Lost Customers Revisiting	44	67	97	187	319	468	160	196	368	516	717	863

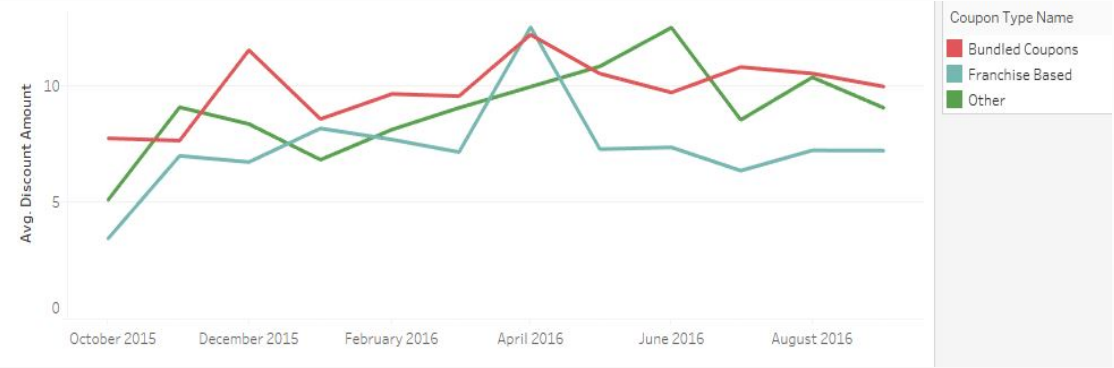
Cluster 1: The Hopper

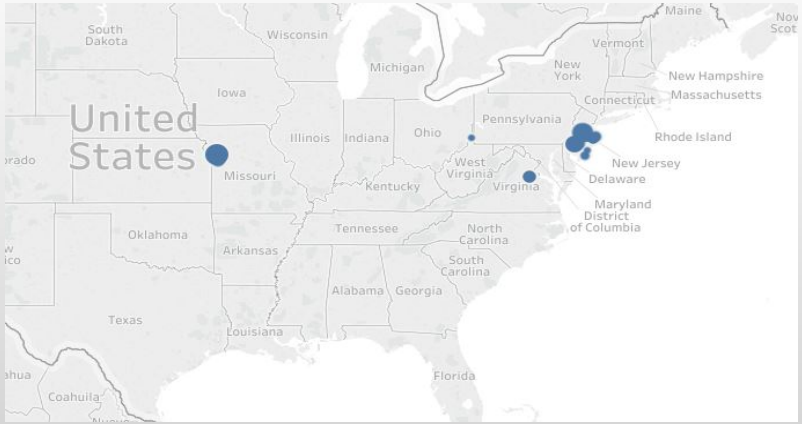
- July-August has the highest Conversion rate of customers from Rejuvenated
- Percentage of conversion from Frequent to rejuvenate is very low
- These customers are looking for Bundled coupons with a higher discount amounts and majority are from Specified areas in **Missouri, Minnesota, Michigan**
- **Recommendation**: Focus needs to be on promoting Bundled Coupons for Customers visiting the stores in these areas





	Oct-15	Nov-15	Dec-15	Jan-16	Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16	Aug-16	Sep-16
Rejuvenated OL	28	80	151	582	949	1,798	23	89	46	768	2,155	1,672
Frequent OL	24	18	23	407	517	751	1	26	18	143	308	261
Total Lost Customers Revisiting	40	56	81	136	234	382	114	188	260	293	622	934





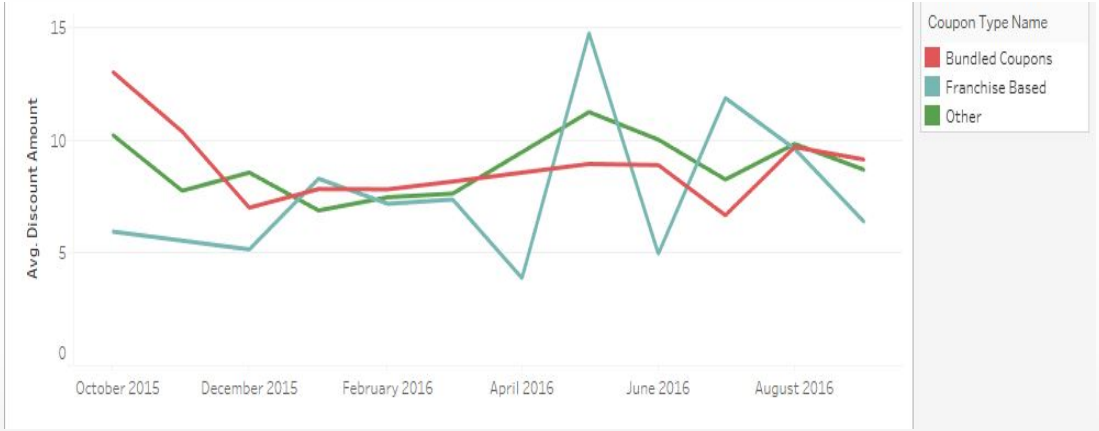
Cluster 2: The Family Guy

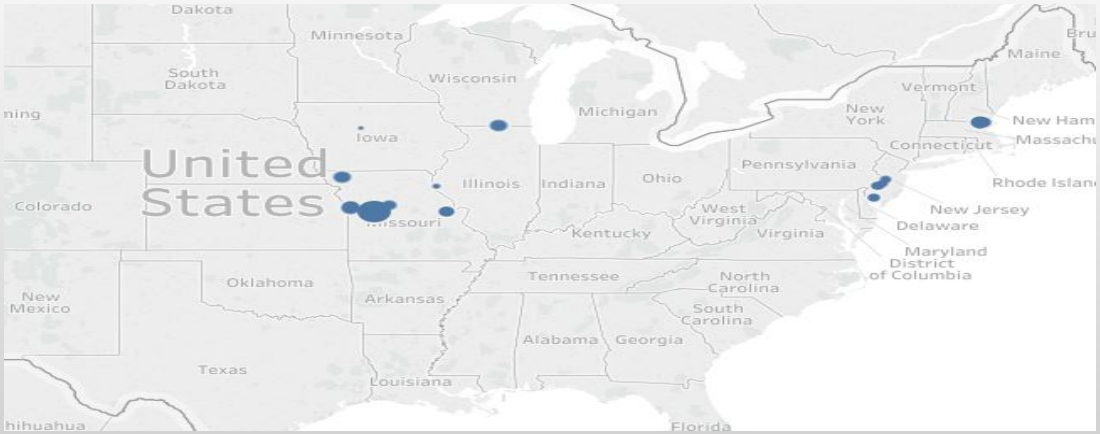
- July – August has the Highest Conversion rate of customers from Rejuvenated
- Percentage of conversion from Frequent to rejuvenate is very low
- These customers are looking for Bundled coupons with a higher discount amounts and are majority from Specified areas in **Missouri, West Virginia, Virginia, New Jersey**
- **Recommendation**: Heavy focus on promoting coupons from March-May and not promoting Franchise based Coupons for customers visiting these stores in these areas.

	Oct-15	Nov-15	Dec-15	Jan-16	Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16	Aug-16	Sep-16
Rejuvenated OL	95	149	195	438	452	734	9	24	33	349	964	1,001
Frequent OL	50	40	49	264	215	381	4	8	9	114	240	203
Total Lost Customers Revisiting	40	118	176	339	467	484	2	26	24	175	457	526

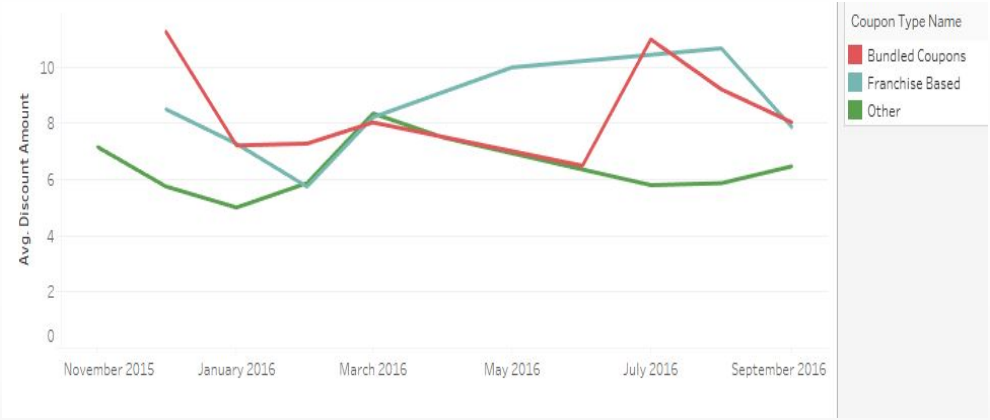
Cluster 3: The Tech Guy

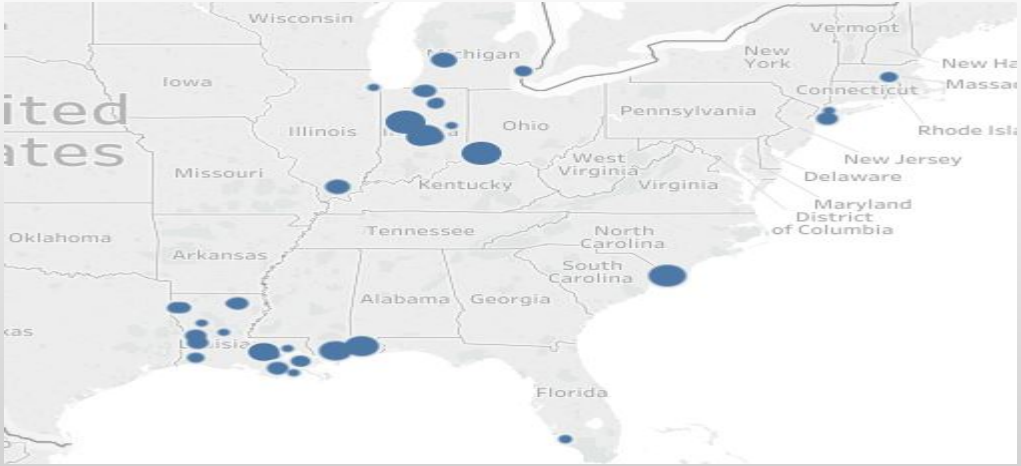
- July-August has the highest Conversion rate of customers from Rejuvenated
- Percentage of conversion from Frequent to rejuvenate is very low
- These customers are looking for Franchise coupons with a higher discount amounts and are majority from Specified areas in **Missouri, Wisconsin, New Jersey, Delaware**
- **Recommendation**: Focus need to be on promoting Franchise and Single Coupons from April- June for customers visiting the store in these areas





	Oct-15	Nov-15	Dec-15	Jan-16	Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16	Aug-16	Sep-16
Rejuvenated OL		9	27	363	277	426	2	13	35	115	140	
Frequent OL		2	2	289	170	197	2	8	5	9	35	
Total Lost Customers Revisiting		5	9	187	265	243	2	2	5	14	44	59





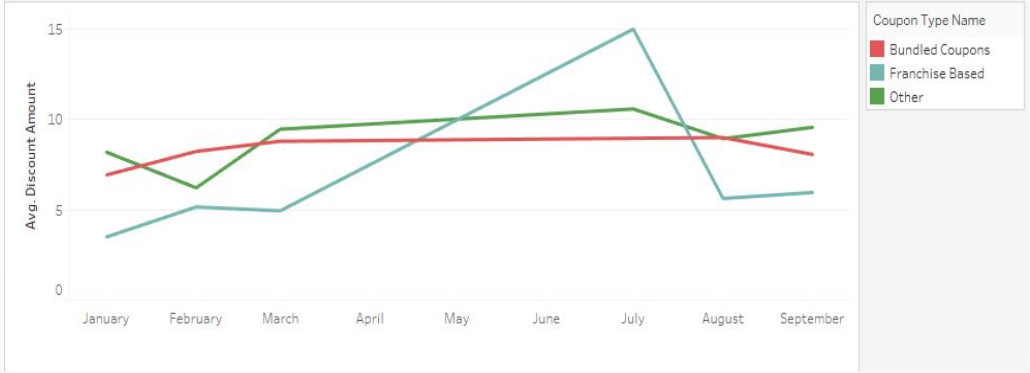
Cluster 4: Pizza Lover

- December-January has the Highest Conversion rate of customers from Rejuvenated
- Percentage of conversion from frequent to rejuvenate is very low
- These customers are looking for Franchise coupons with a higher discount amounts and are majority from Specified areas in **Louisiana, Michigan, Indiana**
- **Recommendation**: Strategic promotion of coupons
- **Single Offers** –From November – December and March
- **Bundled** – January and August
- **Franchise** – March - September

	Oct-15	Nov-15	Dec-15	Jan-16	Feb-16	Mar-16	Apr-16	May-16	Jun-16	Jul-16	Aug-16	Sep-16
Rejuvenated OL				17	75	108				172	203	311
Frequent OL				100	65	127				13	20	19
Total Lost Customers Revisiting				56	102	132				5	25	46

Cluster 5: The Bachelor

- August – September has the highest conversion rate of customers from Rejuvenated
- Percentage of conversion from frequent to rejuvenate is very low.
- These customers are looking for franchise coupons with a higher discount amounts and majority are from Specified areas in Pennsylvania, New York
- **Recommendation**: Heavy focus on promoting Franchise coupons from May – July for Customers visiting the stores in these areas
- Estimated Marketing spend for each customer:
  - $4 for January and February
  - $ 10 for Remaining Months



Objective 2

Coupon Ordering Likelihood

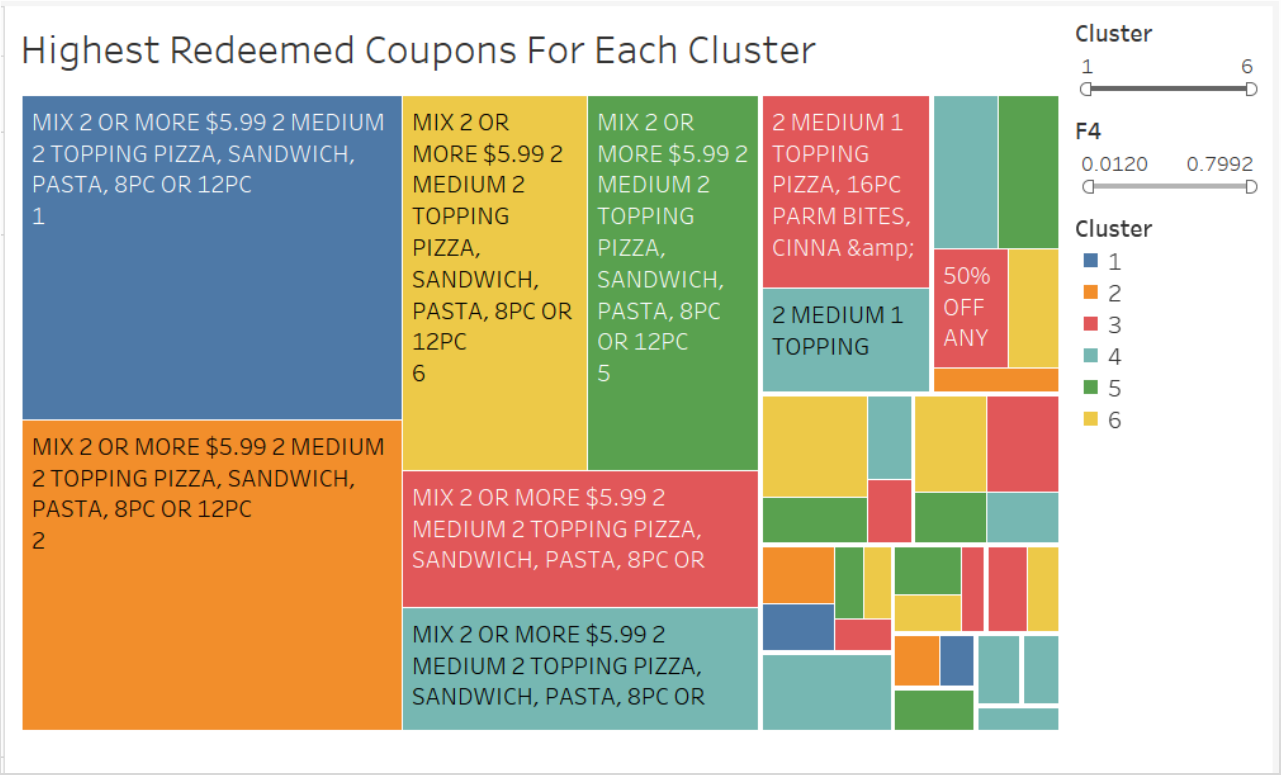
- My goal is to determine which offers would be the most effective to send to each household based on their order history
- Data set used: January'16 – December'16 food chain data containing both demographic data and transactional data pertaining to each purchase
- The idea for this is to determine which offers would be relevant to run and design different variations of offers
- The offer sets are created based on previous orders and offers previously redeemed by the household and by households that have similar demographic profiles
- Aim is to have each household tagged with certain offer sets that would be the most beneficial to mail to them

6 Customer profiles

Created based on customers purchase patterns and transaction history

CLUSTERS	Top 5-10 Coupons %	Coupons used %	Order Amount	Pie Count	Dessert Count	Drink Count	Pasta Count	PURCHASE POWER INCOME	NUMBER OF ADULTS	NUMBER OF CHILDREN	HEAVY INTERNET USER
SANDWICH LOVERS	85%	68%						Moderate to Low (~$50k)	Moderate (1-2 adults)	Moderate (1 or no child)	Moderate User
PASTA CRAVING	85%	68%					**HIGH**	Moderate to Low (~$50k)	Moderate (1-2 adults)	Moderate (1 or no child)	Low-Moderate
SWEET TOOTH	70%	61%			**HIGH**			Moderate (~$60k)	High (>=2 adults)	High (>= 1 child)	Low-Moderate
SODA LOVERS	65%	69%				**HIGH**		Low (~$40k)	Moderate (1-2 adults)	Moderate (1 or no child)	Low-Moderate
PIZZA LOVERS	61%	63%		**HIGH**				High (>$60k)	High (>=2 adults)	High (>= 1 child)	Low-Moderate
MUNCHIES	66%	72%	**LOWEST**					Moderate to Low (~$50k)	Moderate (1-2 adults)	Moderate (1 or no child)	Low-Moderate

Highest redeemed coupon for each cluster



- If we observe the data, the top coupon for clusters 1 - 4 is common
- For Clusters 3 and 4, the 2nd most used coupon is **2 Medium 1 topping pizza, 16 pc Parm bites Cinna & cooldrink.** This is because they are Dessert Lovers and Soda Lovers.

Recommendation: We could modify the top coupon to include drinks and desserts.

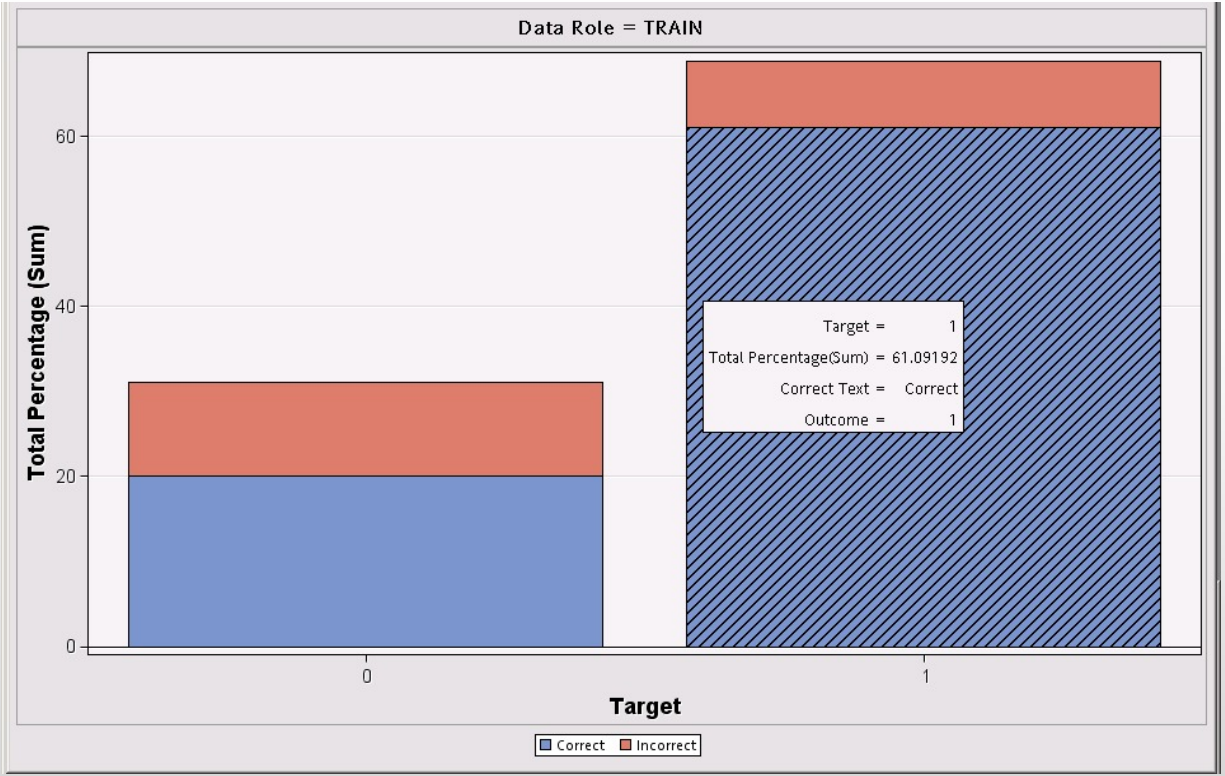
Logit for predicting coupon usage

Dependent Variable – Coupon Usage

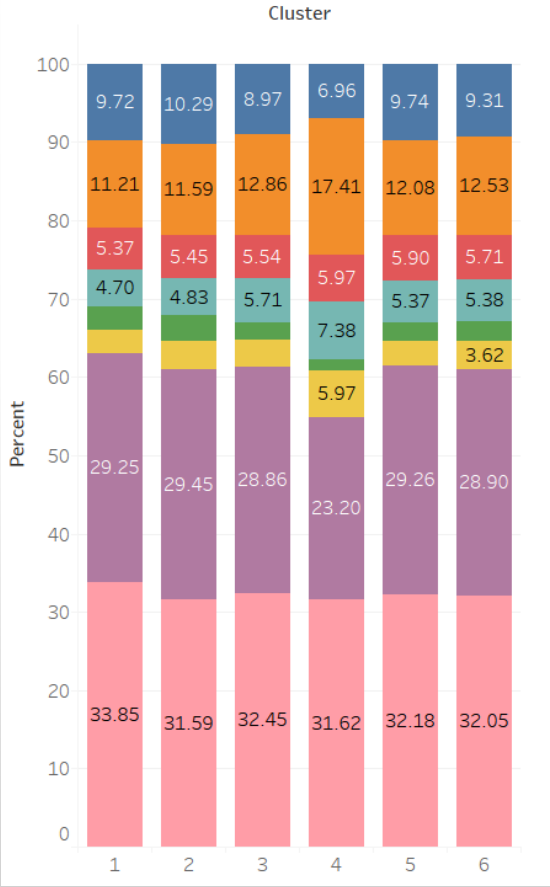
Category : Binary Variable

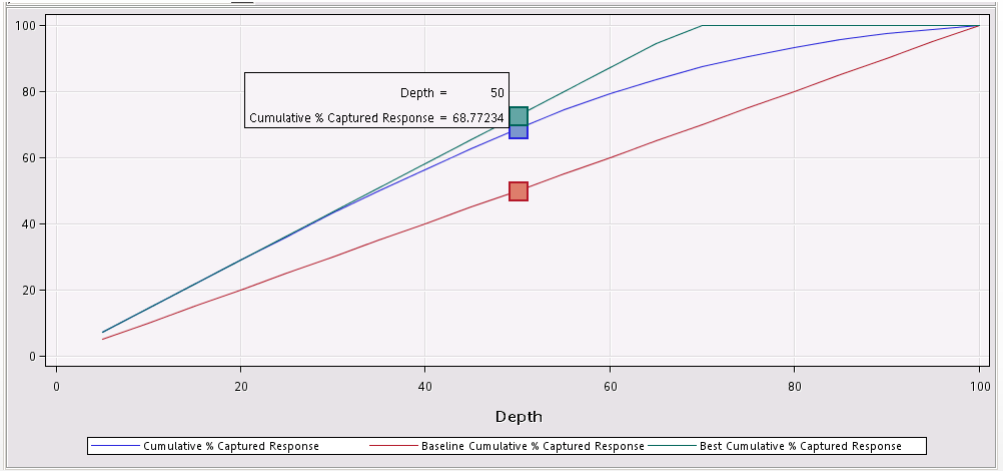
Significant Independent Variables:

- Customer Amount
- Side Order Count
- Chicken Order Count
- Dessert Order Count
- Pasta Order Count
- 2 litre Drink Order Count
- Pie Count
- Sandwich Order Count



Analysis of results from Logit





- We Created the Predicted Probabilities for the likelihood of usage of a coupon.
- we further analysed the First Five deciles which consisted of customers with at least 80% of estimated response and 68% of the total Coupon Users.
- We could observe a high risk customers in dessert lovers which can be avoided by modifying the 9193 coupon like suggested earlier.

Objective 3

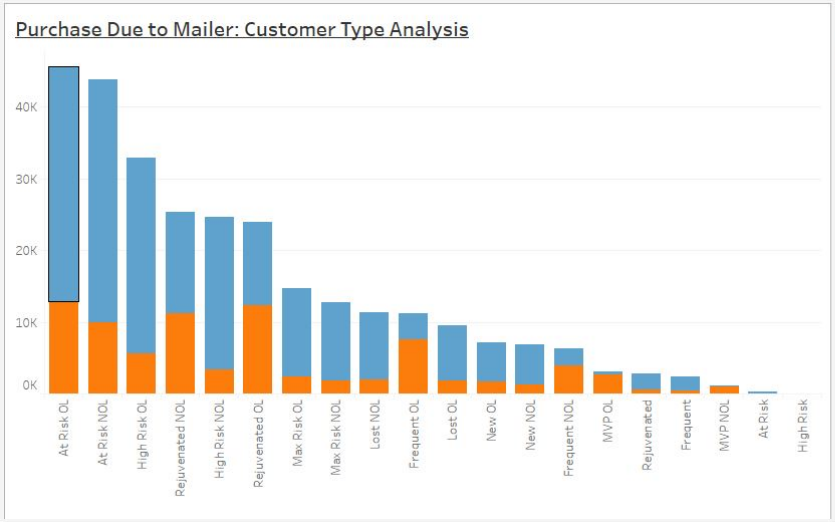
Purchase Ordering Likelihood Due To Mailer

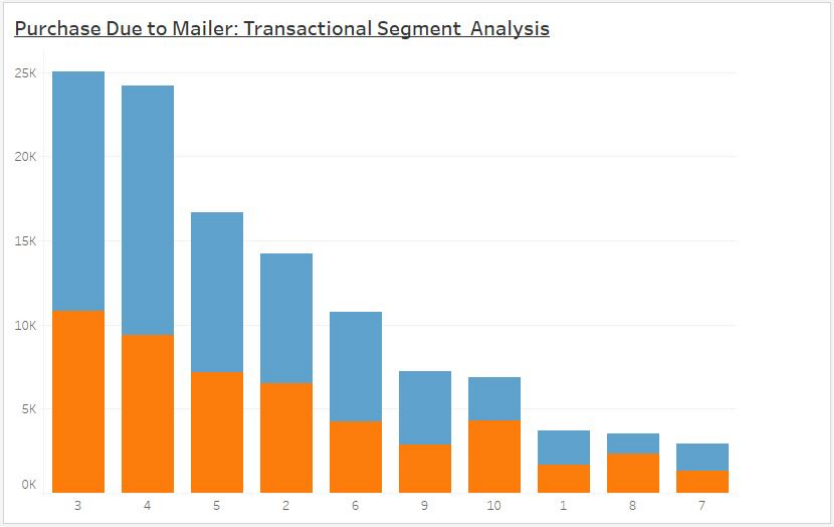
- My goal is to better understand which offers should be promoted to which customers via coupon mailers
- Segmented customers based on their transactional history to determine coupon sets that best fit each segment
- Aim is to devise a method of target selection in order to maximize the return on marketing spend
- Data set used: January – March 2016 set of mailed and control customers with their response information (Purchased within 4 weeks/ Did not purchase within 4 weeks)
- Moved any customer that had no transactional segment into11th segment, Unknown
- Calculated the average number of items ordered per visit for each customer in the previous quarter
- Converted any customer without previous order information to 2.72, the average of all customers average items

10 Customer profiles

Created based on transactional history

CLUSTERS	CUSTOMER PROFILES
Food Aficionado	Tends to get a large specialty pie or a large with multiple toppings. Generally order for dinner.
Soda Lover, Light on Food	2 top hand tossed medium side order about 1/2 time dinner or late night, not heavy on lunch orders. Tend to order soda.
Passive Customer	Avg. 1 pizza slightly below ½ time for side orders, usually medium or large, trend towards lunch orders over dinner orders.
Large Pizza for Dinner	1-2 large pizzas often on Friday night for dinner
2 Medium Sized Pies	2 pizzas generally medium
Side Bread Consumer	Tend to order med pizza and bread side. just about always order a side generally order for dinner
Pizza only With Chicken	2 med pizza generally hand tossed, always order a side, generally chicken order at dinner time
Readymade Meal Seeker	1 hand tossed medium pie, frequent orderers, order during work week
Chicken Hunter	Order for chicken, pizza is an afterthought
Weekend Pizza Party	Frequent weekend orderers, 1 medium order sides about 1/2 time almost even mix of what they order for the side primarily dinner orders
Segment Unknown	This group is all customers whose transactional segment is unknown





Purchase due to mailer by group membership

- Customer Type is the recent segmentation already in use
- Transactional Segments are shown as the segmentation number defined earlier
- Segment 11, or Unknown, has been excluded from current graph. It had 170,270 members total and 31,838 purchases for a rate of 18.7%.

Model – Logistic Regression

- The following model was transformed using logistic regression to find each customer's individual probability of purchasing within 4 weeks of receiving a mailer:

Purchase = Customer Type
+ Transactional Segment
+ Average Total Items
+ Number of Adults
+ Number of Children

Exploratory Test & Goodness of Fit

- My model was created using only customers who were sent a mailer.
- Wanted to understand the effect of receiving a mailer vs not receiving a mailer.
- Created a second model using the same variables, adding mailed as a variable to test this effect.
- The odds ratio of this estimate was 3.128, indicating that receiving a mailer has a strong effect of the predicted probability
- Model Fit Statistics indicated that the model is stronger than both

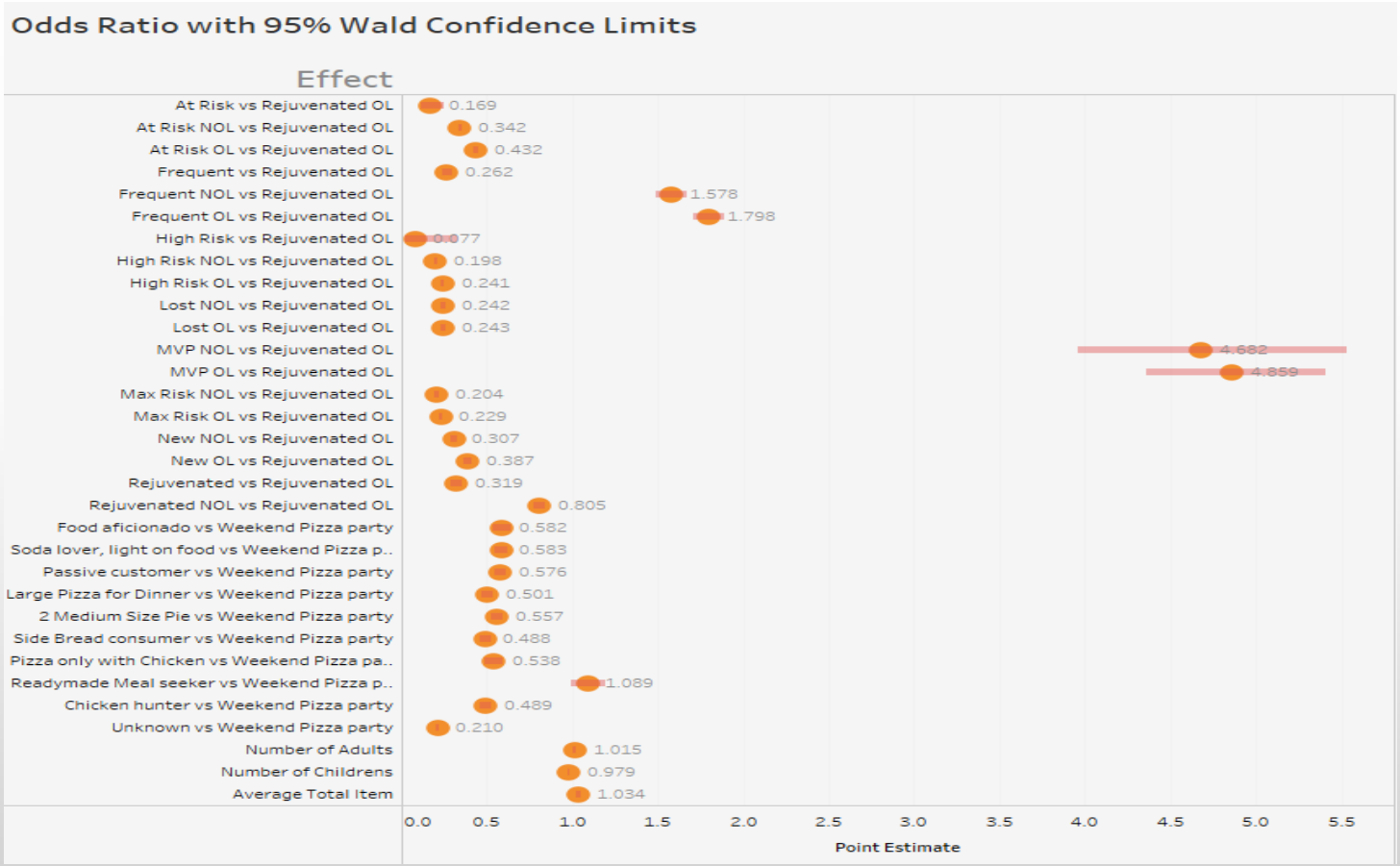
  A lack of model

  &

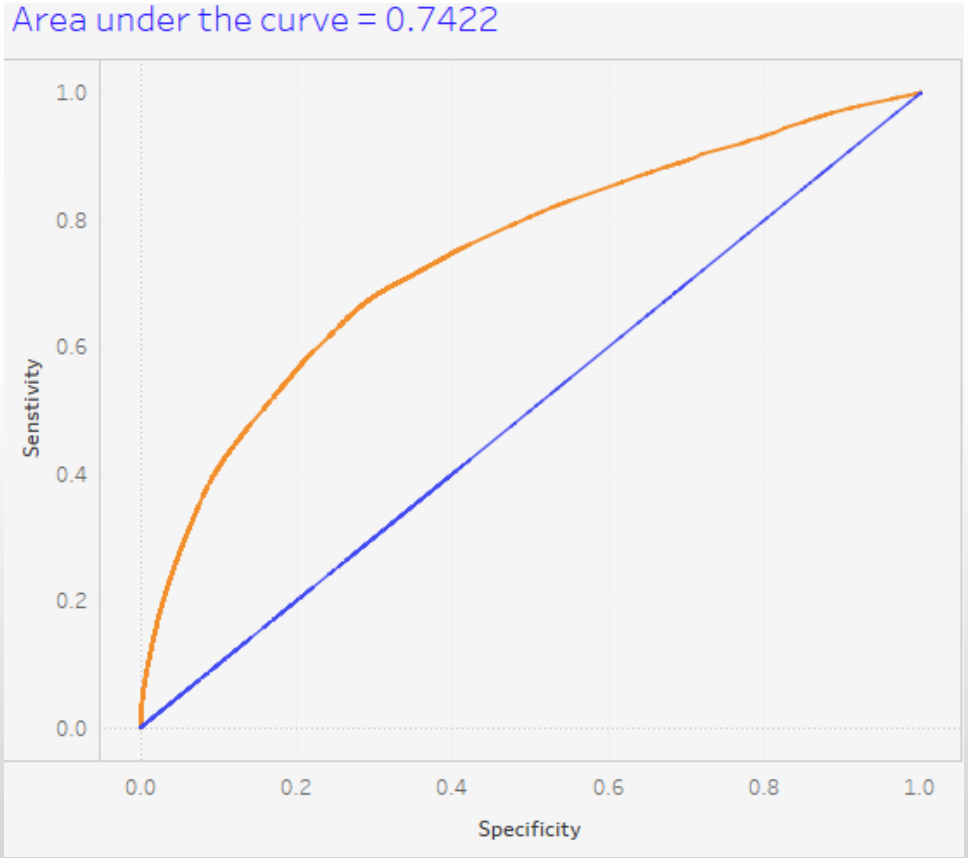
  A model created from only Customer type

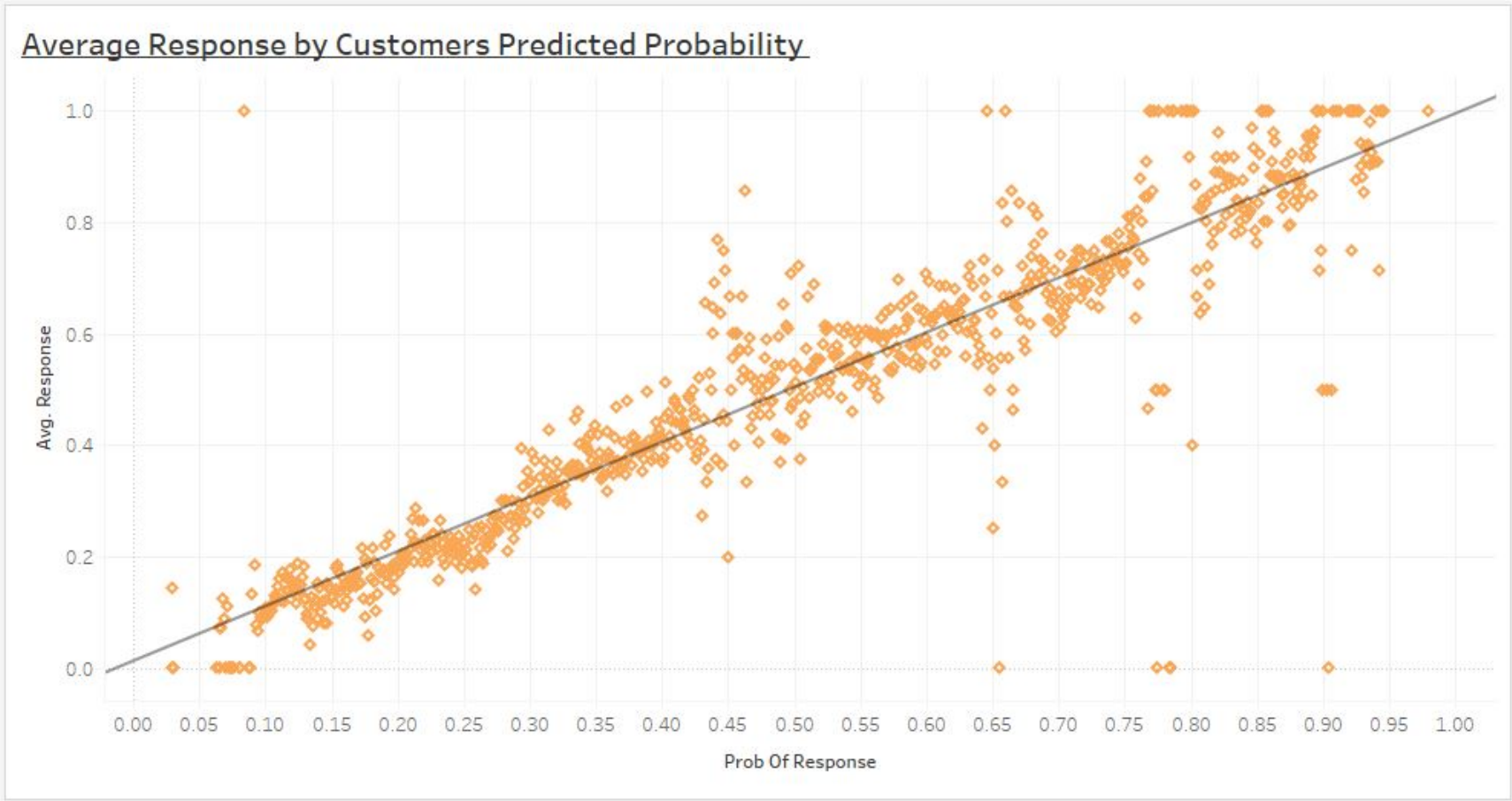
- All variables were found to be significant at $P < 0.01$. 33 of 34 point estimates were found to be significant at $P < 0.01$, 1 point estimate was significant at $P < 0.10$

Odds Ratio Point Estimates with Conf. Intervals

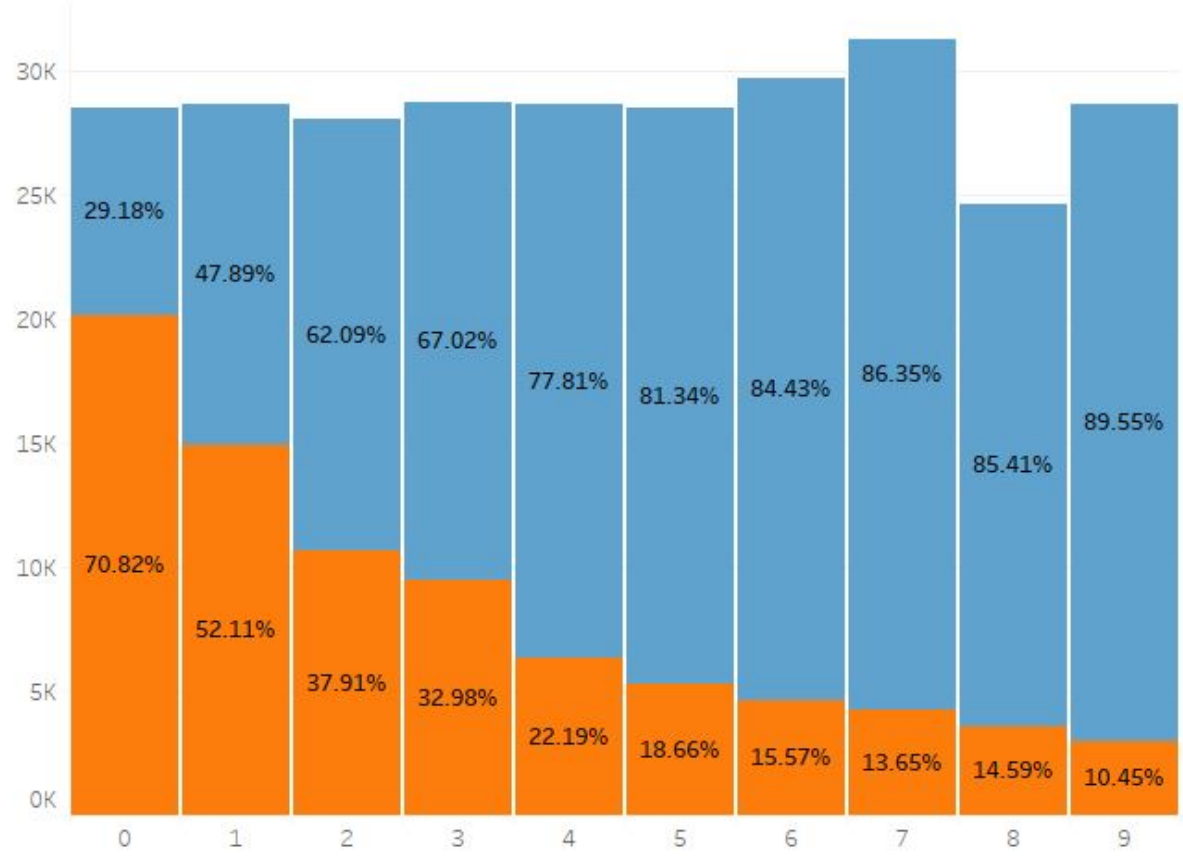


ROC Curve


Area under the curve = 0.7422
1.0
0.8
0.6
0.4
0.2
0.0
Senstivity
0.0
0.2
0.4
0.6
0.8
1.0
Specificity

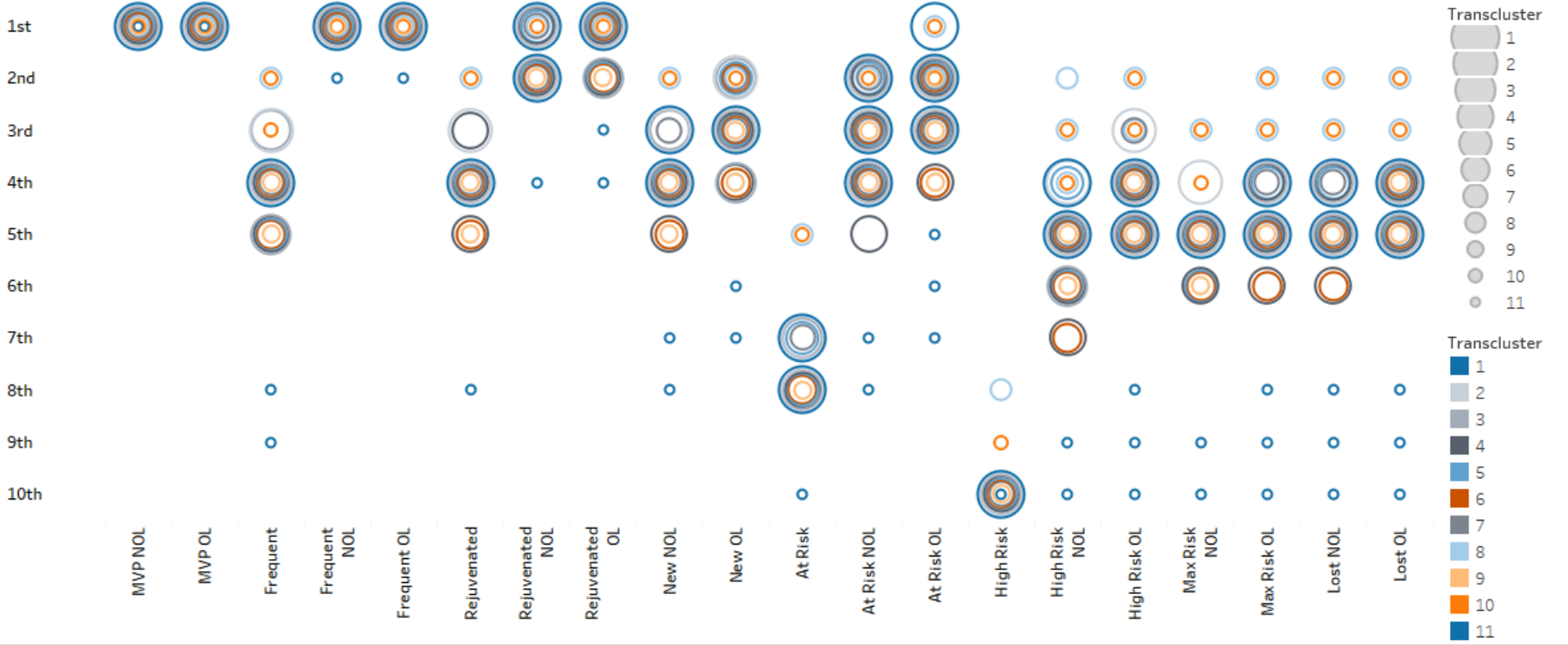

Average Response by Customers Predicted Probability
Avg. Response
1.0
0.8
0.6
0.4
0.2
0.0
0.00 0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95 1.00
Prob Of Response

Purchase Due to Mailer: Predicted Probability Decile Analysis

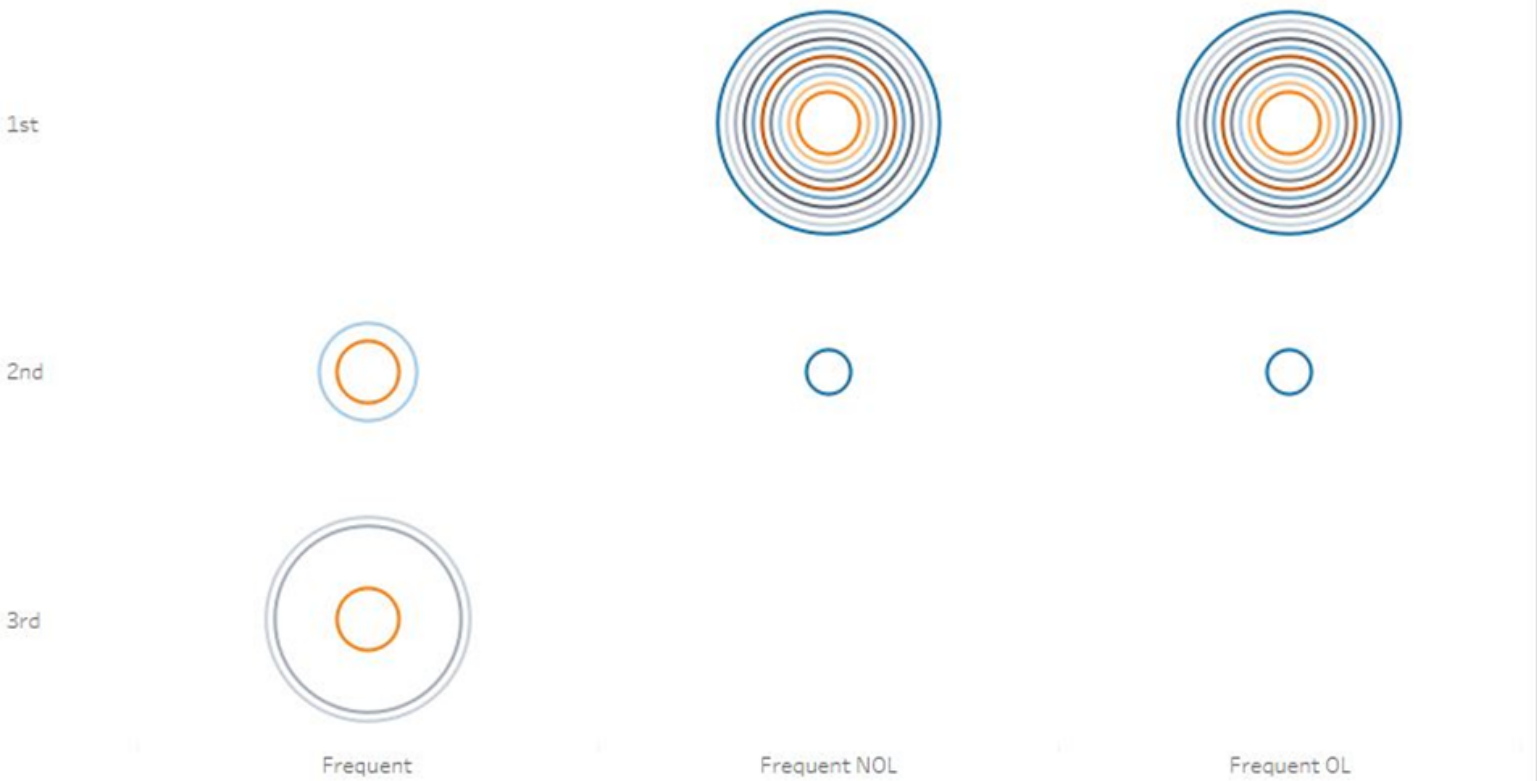


Response
0
1

Decile Ranking Makeup: Customer Type and Segmentation



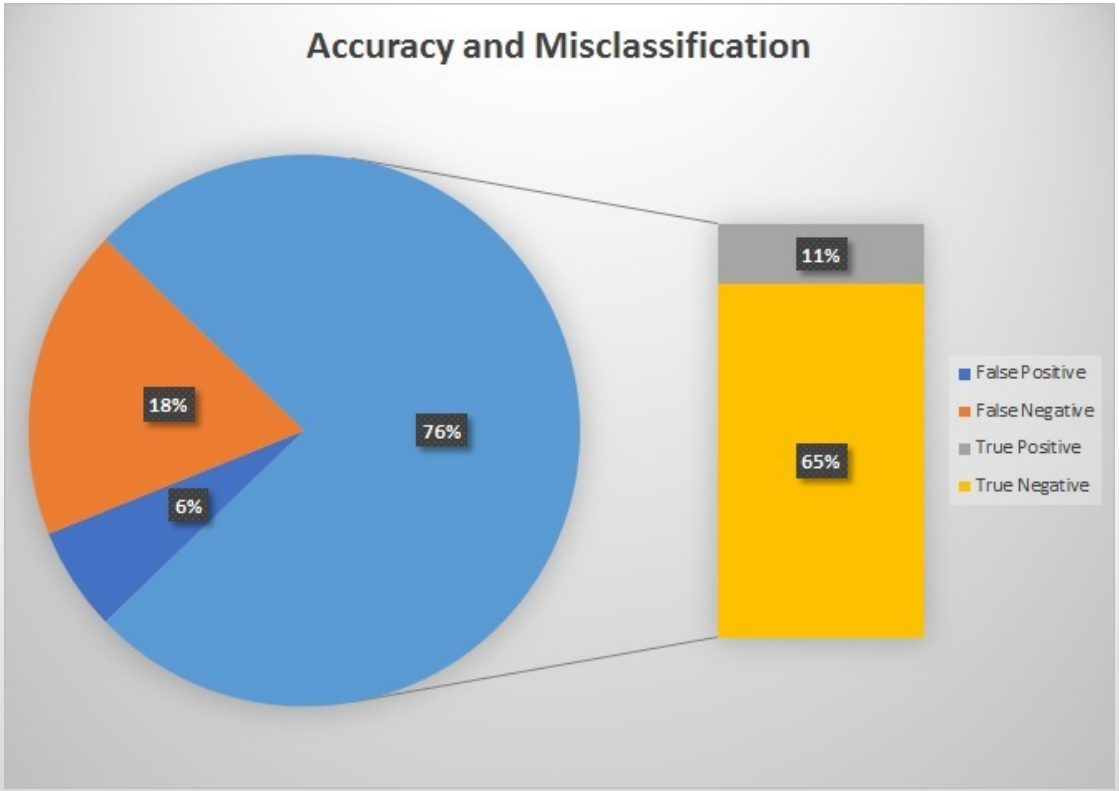
Decile Ranking Makeup: Customer Type and Segmentation



Transcluster
1
2
3
4
5
6
7
8
9
10
11

Transcluster
1
2
3
4
5
6
7
8
9
10
11

Accuracy and Misclassification



False Negative rate is currently higher than the True Positive

56.96% of False Negatives are in the Unknown segment

Probability cutoff for this test was set to .50

Offer Sets By Transactional Cluster:

Name	1- Food Aficionado	2- Soda Lover, light on food	3- Passive Customer	4 -Large Pizza For Dinner	5- 2 Medium Size Pies	6- Side Bread Consumer	7- Pizza only with Chicken	8- Readymade Meal Seeker	9- Chicken Hunter	10- Weekend Pizza Party
Short Description	Orders irregularly but get large specialty pizzas and/or lots of toppings, usually for dinner	2 topping hand tossed medium, side order about 1/2 time dinner or late night, not heavy on lunch orders	Avg 1 pizza, usually medium or large, side orders about 45 percent of time slight trend towards lunch orders over dinner orders	1-2 large pizzas for dinner.	2 pizzas generally medium	Tend to order med or large pizza and bread side, order a side ~98% of time, generally dinner orders	2 med pizza generally hand tossed, always order a side, generally chicken order at dinner time	1 hand tossed medium pie, frequent orderers, side 1/2 time, drink ¼ time, tend to order during work week	Order for chicken – 2 chicken per order is average, may order a Med or a Large as well. Drinks ¼ time	Frequent weekend orderers, 1 medium order sides about 1/2 time almost even mix of what they order for the side primarily dinner orders
Offer1	Mix 2	Mix 2	Mix 2	Mix 2	Mix 2	Mix 2	Mix 2	Mix 2	Mix 2	Mix 2
Offer2	2 Large 3 toppings	2 Med 2 toppings and a 2 liter	Mon-Thurs Carryout Large 3 Topping	Mon-Thurs Carryout Large 3 Topping	Med Pan 2 toppings	Large 3 topping and Bread Side	Med 3 toppings, chicken, 2 liter	Mon-Thurs Carryout Large 3 Topping	2 chicken & Lg 2 topping	Med Pan w/ 2 toppings
Offer3	Large specialty Pizza	2 medium 3 toppings	1 Large 2 Topping	2 Large 2 Toppings	2 Med 3 toppings	Med pizza w/ 2 toppings, bread side and a 2 liter	Med 1 Topping, chicken, 2 liter	Med pizza w/ 2 toppings, bread side and a 2 liter	Lg 1 topping, 2 chicken, 2 liter	Med pizza w/ 2 toppings, side and a 2 liter
Offer4	Mon-Thurs Carryout Lg 3 Topping	Mon-Thurs Carryout Lg 3 Topping	1 Medium 2 topping with a drink offered at lunch time	50% off any pizza at menu price	Med 3 Topping	Med 3 topping, 2 Bread sides	Buy 2 Medium pizzas get 1 chicken free	Med 3 toppings, chicken, 2 liter	Med 3 topping, Chicken, 2 liter	Med pizza w/3 toppings
Offer5	Buy Large Specialty & get Large 2 topping free	Large 3 toppings with 2 liter	2 Large 2 Toppings	2 large 3 topping	2 Med Pan 3 toppings	Buy 2 Medium pizzas get bread free		Med 3 toppings w/ 2 liter		

RECOMMENDATIONS

When planning a general mailer campaign, I recommend targeting those customers falling within the top 2 deciles

When planning a focused mailer campaign, I recommend sending mailers to those customers at higher levels of predicted probability of ordering for the target set

The customers in the Unknown Cluster should be mailed a mix of coupons until their cluster has been determined

Once your target customers are selected, use their Transactional Segment to determine which offer set to mail

ADDITIONAL RECOMMENDATIONS ON THE GO..

Increased sales of add-on items. Giving customers a "Complete experience of the menu"

Pizza Tracker and 3-D modelling feature to attract "Younger users"

Located "over performing stores" by geospatial systems to target ethnic eating habits