inteQ™

AIC Internship Presentation

Doug MacClure

10.13.17

Introduction

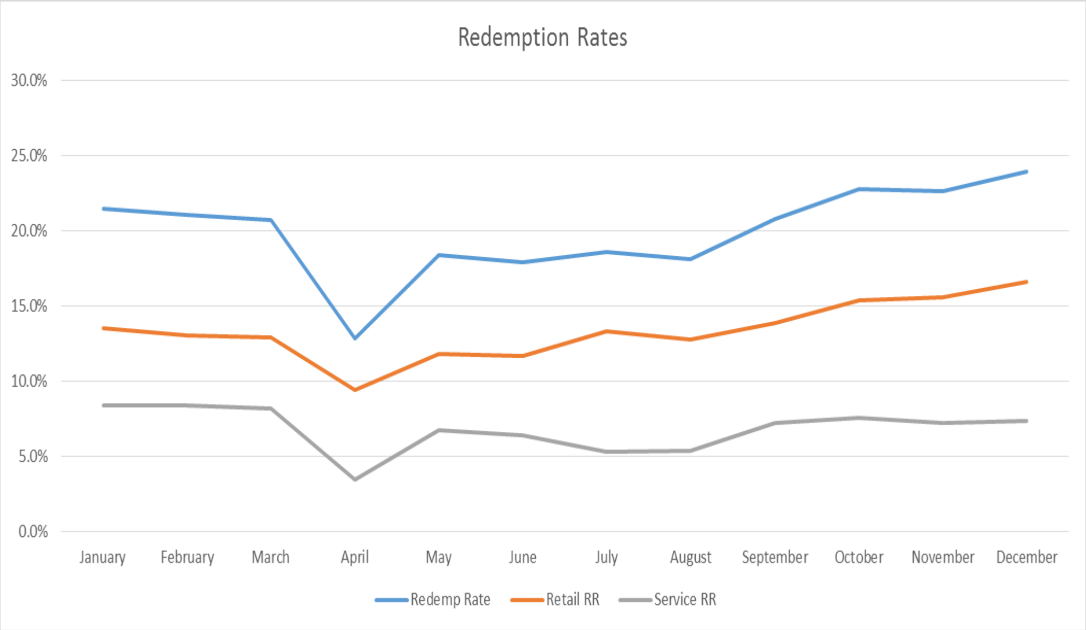
- Inte Q is a marketing research firm, providing analytics to service and retail clients
- The primary service Inte Q offers to clients is the management of customer rewards programs
- To accomplish this, Inte Q houses large data “warehouses” provided by the client
- This data is analyzed primarily with the SAS statistical software package, and organized using SQL calls in SAS

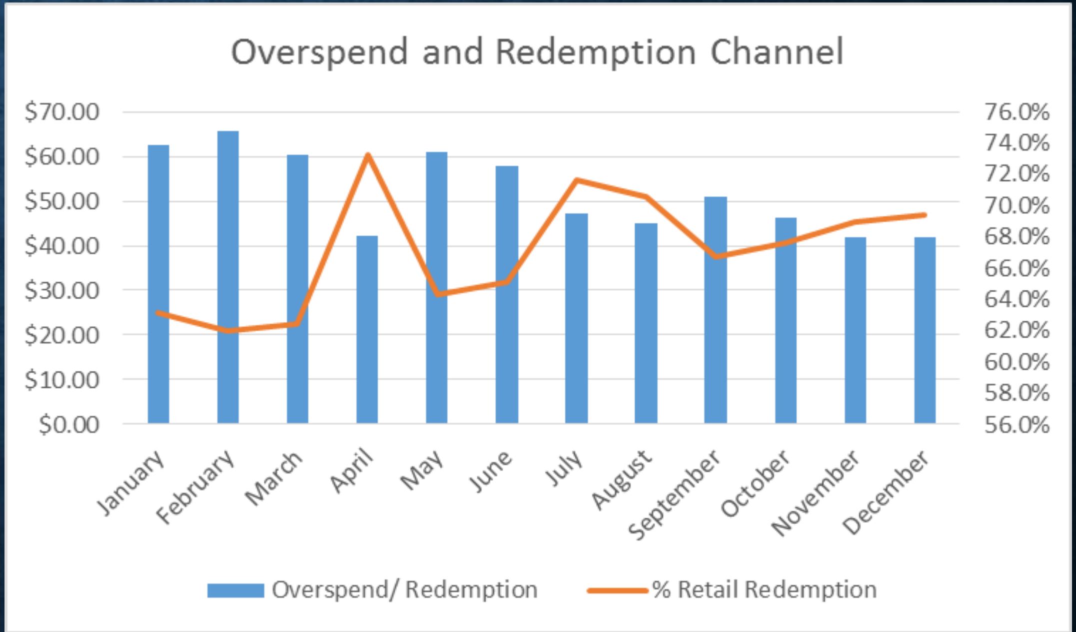
Overview

- Getting started: An analysis of a change in points threshold for a client rewards program
    - Provided the impetus for the Reactivation Model
- Building the Reactivation Model
- Implementation of the Reactivation Model
- Takeaways

Getting Started

- The first phase of my internship primarily involved learning the tools used throughout the internship and the data involved in the Reactivation Model
    - My first task: learn the basics of SAS & SQL and update the data for the Retail Multiplier Analysis report
- In April 2016, an automotive client lowered the $10 reward point threshold from 250 to 100 points to reward retail customers
    - This was one of the client's busiest months in retail reward issuance but saw the lowest average overspend and redemption rate, since members with little to no activity in the previous 9 months were issued a reward





Where to begin?

- Among those with less than 250 points who were issued a $10 reward under the 100 point threshold, over 500,000 members were "lapsing" members
    - A member is considered lapsing if they have made no transactions with the client within the past 9 months
    - These are members with points nearing expiration (13 month window)
    - The $10 reward expires after 3 months
- This large population of lapsing members who were issued a reward provides a wealth of data for a statistical model which predicts the characteristics of lapsing members that are most likely to redeem a $10 reward
    - Such a model can provide a blueprint for turning lapsing members with sufficient activity 9-13 months prior into re-engaged customers

Where to begin?

- However, enticing a lapsing member into simply redeeming a reward is not always profitable, given the cost of the $10 reward and the cost of mailing
    - We want to instead predict the amount a member will spend over a $10 reward (overspend)
    - Those that do not redeem the $10 reward will be considered to have $0 overspend
- A model which predicts the value of a continuous variable for each lapsing member is required
- Variety of models to choose from: Ordinary Least Squares linear regression, decision trees, other high-powered statistical models

Where to begin?

- Issues to consider:
  - People are not entirely rational in their purchases
  - Limitations of capturing personal data beyond transactional data
  - Variance vs. consistency
  - Capturing the signal in the noise
  - Over-fitting vs. Under-fitting the data
  - Implementation of model in practice has low risk of causing negative effects
    - If a lapsing member doesn’t redeem the reward, the costs are marginal
  - The model will have repeated application, and should be easy to understand
    - The client’s upper management will need to understand the model

Where to begin?

- The Reactivation Model is a multi-variate linear model which determines the predicted overspend of a lapsing member
- Since overspend has continuous data-type, Ordinary Least Squares regression is used to compute the coefficients of the linear model

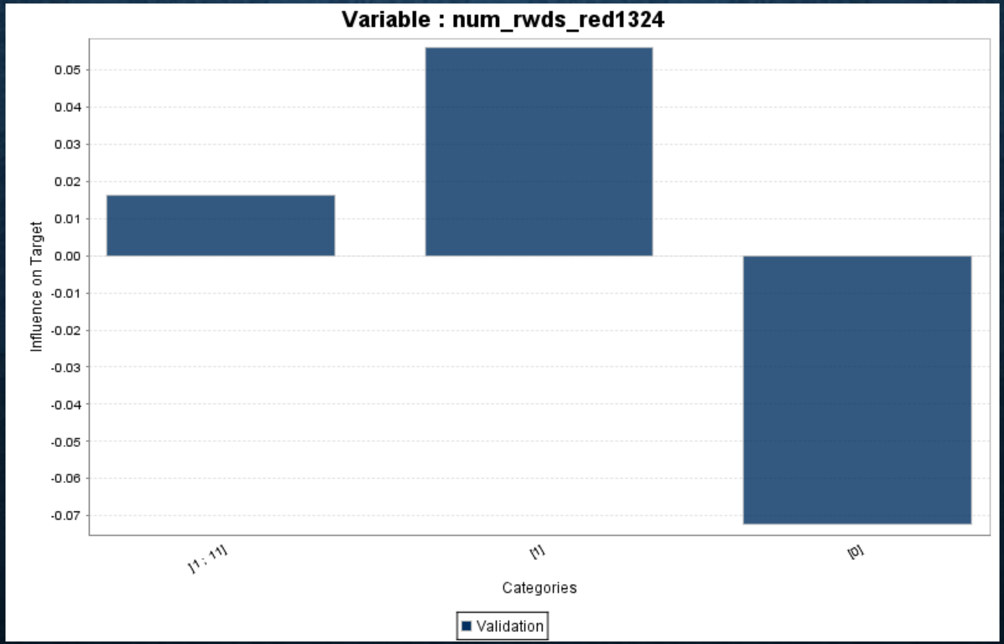
Building the Reactivation Model

Predictive Variables

- Predictive variables for models which predict customer behavior usually include RFM (recency, frequency, monetary) metrics
- It is also necessary to become familiar with the client's business
  - The client is an automotive retail and service company, hence it is necessary to create predictive variables which differentiate between retail and service transactions
  - Since most members' transaction activity varies from year to year, we consider 0-12 month and 13-24 month increments of member transaction activity
- Some important metrics:
  - 13-24 month net spend
  - Shop in Department X 13-24 months prior flag
  - Number of transactions 13-24 months prior
  - Number of rewards earned 13-24 months prior
  - Tenure in rewards program
  - Number of service transactions 0-12 months and 13-24 months prior
- In the end, almost 250 predictive variables were considered for this model

Model Reduction

- Next step: remove highly correlated variables and variables with low significance towards predicting overspend
- Use A.I. for model reduction: SAP's KXEN statistical software package
- KXEN does not provide the best possible list of predictive variables
    - Issue: over-bucketing of variables can lead to problems
- Use KXEN to reduce the model in multiple ways, removing different correlated variables at each step
- After several iterations, end up with a much smaller set of predictive variables (final model used less than 10 predictive variables)



OLS Regression

- With a reduced model (less than 20 variables), use SAS's OLS Regression procedure to develop a linear multivariate map which predicts a member's overspend

```
proc reg data = modeling_dataset_1 outest = estimates;
        model totoversp =
        num_rwds_red1324 depart_98_f11324 depart_98_f10012
        NetSales_1324 num_units_bought0012
        days_in_program_1664to2506 days_in_program_0to1663
        No_of_Visits_S_0012_2to47 No_of_Visits_S_0012_1
        No_of_Visits_S_0012_0               / &mdl_options;
    output out = modeling_dataset_1 p=yhat;
run;
```

- Use a stepwise selection process, where variables are selected for the model one-by-one if a prescribed significance towards predicting overspend is reached
    - A predictor's significance towards overspend is computed using the R-squared statistic
    - At each stage, predictors which entered the model are also considered for removal
        - Two predictors may be correlated – if the currently evaluated predictor has more significance towards predicting overspend, the other predictor is removed from the model
- Perform the regression using the training data set (70% of the population), and test the efficacy of the model on the validation data set (remaining 30% of population)

OLS Regression

- After running the proc reg call in SAS, we get a list of variables and coefficients
- Example model (not the final model):

```
score=1.53172 + 6.8783*rwd_red1324fl_1 + 1.10465*depart_98_fl1324 + 1.32325*depart_98_fl0012 + 0.32254*NetSales_1324_1
        + 0.38185*num_units_bought0012_1 + 1.4741*days_in_program2507to9999 - 0.45768*days_in_program_0to1663;
```

- Rank the validation dataset by predicted overspend (score) into deciles (increments of 10%)
- Evaluate the efficacy of the model in predicting overspend by computing the model lift (Cumulative Average Overspend up to current decile / Cumulative Average overspend for all deciles)

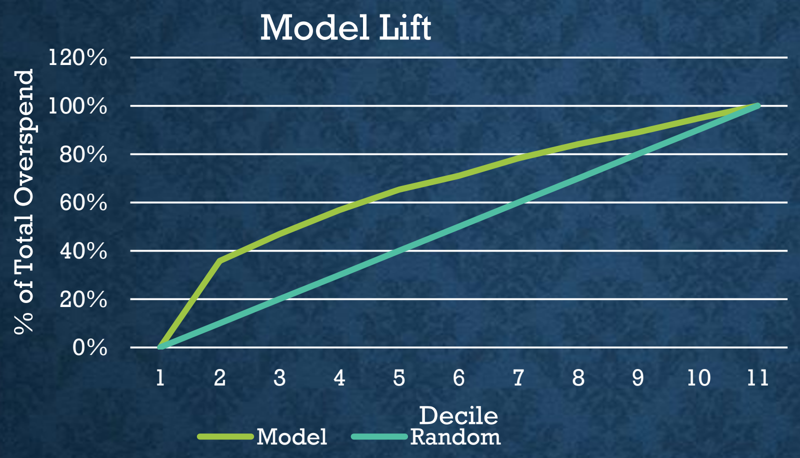
Decile	Number of Members	Predicted Overspend	Actual Overspend	Cumulative Actual Overspend	Cumulative Average	Model Lift	Cumulative % of Overspend	Sales	Response Rate
1	14769	$95,896	$104,028	$104,028	$7.04	2.8	36%	10%	14.5%
2	14770	$54,889	$54,722	$158,751	$5.37	2.1	47%	20%	7.1%
3	14769	$43,392	$37,894	$196,645	$4.44	1.8	57%	30%	7.7%
4	14770	$37,117	$31,416	$228,061	$3.86	1.5	65%	40%	6.8%
5	14726	$30,842	$34,149	$262,210	$3.55	1.4	71%	50%	5.8%
6	14814	$27,809	$23,057	$285,266	$3.22	1.3	78%	60%	6.1%
7	14778	$25,151	$24,814	$310,080	$3.00	1.2	84%	70%	6.8%
8	14946	$20,927	$25,422	$335,502	$2.84	1.1	89%	80%	5.5%
9	14722	$15,761	$18,983	$354,485	$2.66	1.1	95%	90%	4.9%
10	14631	$13,578	$18,169	$372,655	$2.52	1.0	100%	100%	5.1%

Evaluating a candidate model

- Next, plot Cumulative % of Overspend and Random Assignment for each decile
    - A good model will produce a model lift with maximal area between the two curves and be approximately concave down
    - First time through: issues with model consistency – too many bucketed/discrete variables

Evaluating a candidate model

- After several attempts and changes, we arrive at an acceptable model



Decile	No. Of Members	Predicted Min Overspend	Predicted Max Overspend	Actual Average Overspend	Culm. Ave. Overspend	Culm. Predicted Overspend	Culm. Actual Overspend	Redemption Rate	Cumulative Sales Lift (Validation)
1	49,232	$4.29	$57.81	$6.85	$6.85	$319,649	$337,209	14.8%	2.8
2	49,262	$3.19	$4.29	$3.49	$5.17	$502,543	$508,954	7.3%	2.1
3	49,202	$2.74	$3.19	$2.73	$4.36	$646,931	$643,409	7.8%	1.8
4	49,231	$2.23	$2.74	$2.39	$3.86	$770,584	$760,922	6.8%	1.5
5	49,234	$1.98	$2.23	$2.14	$3.52	$873,550	$866,414	5.9%	1.4
6	49,230	$1.78	$1.98	$1.71	$3.22	$965,888	$950,647	6.1%	1.3
7	49,232	$1.62	$1.78	$1.65	$2.99	$1,049,583	$1,031,993	7.0%	1.2
8	49,663	$1.15	$1.62	$1.55	$2.81	$1,119,023	$1,109,021	5.9%	1.1
9	49,263	$1.01	$1.15	$1.29	$2.64	$1,171,750	$1,172,380	5.0%	1.1
10	48,769	-$1.01	$1.01	$1.06	$2.49	$1,216,993	$1,224,194	5.1%	1.0

Implementation of the Reactivation Model

Implementation of the Model

- The Reactivation Model showed that the top 30% of the file had the top 3 redemption rates among other deciles, and contributed to 57% of the total overspend, with an overall redemption rate over 10%
    - Retail redemption rates were under 10% and service redemption rates were under 5% in April 2016
    - Redemption rates for lapsed members in April 2016 was around 7%
- The Reactivation Model will be used twice annually (6 months apart) to target members that have not shopped with the client in the past 9-13 months and re-engage these members with the client
- The Reactivation Model will be used in addition to standard monthly rewards sweeps

Profitability

- Using the model to mail lapsing members with at least 100 points a $10 reward, the client can expect a 10% redemption rate, and $4.36 in average overspend among all lapsing members (including those that do not redeem the reward)

- Considering the cost of mailing the reward and the cost of the $10 in-store credit, using the Reactivation model is marginally profitable

- The value in using the Reactivation model is to entice lapsing members into migrating towards more profitable (high transaction volume and/or high spending) segments

What did I learn?

- Statistical Tools:
  - SAS, SQL, KXEN
- Presentation Tools:
  - Excel, PowerPoint
- OLS Regression & Model Building
  - Data and metrics are meaningless without understanding the question you are trying to answer
- Determining whether an enrollment/sales campaign was successful
  - How is it determined that an event was successful on a particular day? What is the store's performance compared to?
- Marketing
  - Sitting in on calls and meetings

Takeaways

- What I liked
  - Work environment
  - Great coffee
  - My own desk/office
  - Good pace of work, starting with a learning focus and ending with a production focus