Capital One Variable Creation

October 30, 2013

1 Variable Ideas

1.1 General Customer Info

- Proportion of purchases customer makes online ($p \in [0,1]$) [Andy Says: `SQL code scripted in VarCreation.R`]
- Customer online transaction count for each industry name (discrete, numeric) – might purchase shoes online but not gasoline or groceries.
- Customer transaction count for each industry name (discrete, numeric) – might not have a car and not need to purchase gas. [Andy Says: `SQL code scripted in VarCreation.R`]
- Total transactions (discrete, numeric) [Andy Says: `SQL code scripted in VarCreation.R`] [Andy Says: `Important as Build data sets only contain high frequency users`].
- Home zip code (character) (also used for customer-merchant specifics).
- Shopping zip codes (list - character) (zip codes where brick-and-mortar purchases are made).
- Transactions in last XX months (discrete, numeric) (identify inactive cards/users) [Andy Says: `SQL code scripted in VarCreation.R for last 3 months`]
- whether they shop at one zipcode only (maybe 2-3 neighboring zipcode). (this can be used to cluster customers (mobility)) [Andy Says: `SQL Code scripted in VarCreation.R`]
- Number of merchants customer shopped at [Andy Says: `SQL Code scripted in VarCreation.R`]

1.2 Merchant Info - intermediate step

- Merchant Zip Codes for all locations

1.3 Customer-Merchant Specifics

1. Frequency of shopping at Merchant
2. Distance to merchant from customer using zip centroids ($\mathbb{R}_+$).
3. Mechant and customer zipcode match (0-1 binary)
4. Merchant and customer shopping zip code match (0-1 binary)
5. Shop at competitor (same industry)

6. shop at competitor in same zip (identify grocery, restaurant, gas patterns, potentially not likely to shop at competitor without coupon) [Andy Says: Could also make a good story in the write-up, bill it as a way for merchants to identify their competition's customers and issue coupons to them]